01: Inaugural Issue

asterisk

Jared Leibowich *Modeling the End of Monkeypox* / Kelsey Piper *Review: What We Owe The Future* / Christopher Leslie Brown *Making Sense of Moral Change* / Kevin Esvelt *How to Prevent the Next Pandemic* / Stuart Ritchie *Rebuilding After the Replication Crisis* / Dietrich Vollrath *Why Isn't the Whole World Rich?* / Scott Alexander *Is Wine Fake?* / Karson Elmgren *China's Silicon Future* / Fred Kaplan *The Illogic of Nuclear Escalation* / Xander Balwit *They May As Well Grow on Trees: The Future of Genetically Engineered Livestock*

Asterisk
2150 Shattuck Ave Fl 12
Berkeley, CA 94704-1345

Editor in Chief: Clara Collier
Managing Editor: Jake Eaton
Copy Editors: Peter Kranitz, James Hu
Fact Checkers: Dale Brauner,
Matt Mahoney

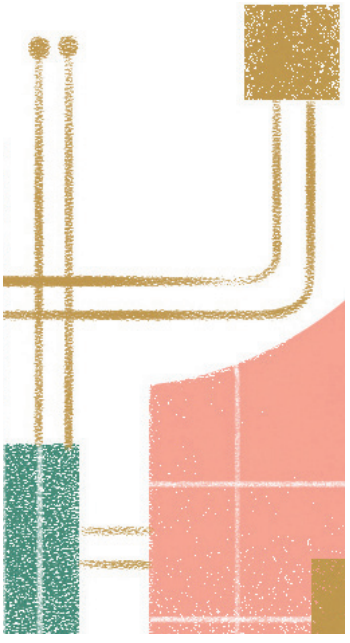
Design: Sarah Gephart/MGMT. design
Web: Marie Otsuka, Minkyoung Kim

Contact: info@asteriskmag.com
asteriskmag.com

Subscriptions: $35/year (general),
$15/year (students).
Contact subscriptions@asteriskmag.com

Asterisk is fiscally sponsored by
Effective Ventures, a 501(c)(3) nonprofit,
and funded by a generous grant from
the Open Philanthropy Project. Special
thanks to the Constellation staff for
their tireless support.

Printed in Canada by Hemlock Printers.



COVER BY
Eleni Debo

This is the first issue of Asterisk, a magazine about the world and what it takes to make sense of it.

We can see our lives getting stranger and harder to predict by the minute, and we're worried that public discourse is failing to rise to the challenge. We're curious about emerging technologies and the people who use them. We think that artificial intelligence is going to change everything — so we're obsessed with AI, and we're obsessed with everything else. We think that we — not our grandchildren — might move to Mars or upload ourselves to the cloud. We're not sure we're going to live to retirement. We are open to the possibilities of a future that sounds like science fiction without losing sight of the messy details that make up reality.

In the end, we don't know what's going to happen next — and we don't think anyone else does either. Our biggest questions don't lend themselves to certainty. Unfortunately, our current media has a deep allergy to being confused in public — which means that the things we need to understand the most are the ones where conversation is shallow, inadequate or simply absent. We don't like being wrong — but, more than that, we don't like being ignorant, so we're going to muddle through the thorniest, most important open problems we can find.

Our writers think critically, probabilistically — and, always, out loud. During the COVID-19 pandemic, we saw too many examples of journalists and public health officials oversimplifying nuanced issues based on how they wanted the public to act — noble lies that for the most part didn't work. Early in the outbreak,

some officials claimed that masks were ineffective in order to save masks for health care workers — but the result was long-standing distrust of masks. When the FDA warned people not to take ivermectin, they called it horse paste rather than saying "the studies showing that it works are small, low-quality and sometimes fraudulent." For a while, data wasn't even collected about post-vaccine infections because of worries that people wouldn't get vaccinated if they knew the vaccines' protection was imperfect.

At Asterisk, we trust our readers. We are honest about what we know and what we don't. We are transparent in our reasoning — you will never be confused about why one of our writers comes to a conclusion. We will be specific enough to sometimes make mistakes, and we'll admit them when they happen. An Asterisk article won't tell you what to think. Instead, it will leave you with a stronger, more complete understanding of the world — because we believe that helping people understand the world leaves them better equipped to change it.

There has never been a time when clear thinking mattered more. We're on the brink of massive technological changes that there's no going back from. We're scared. We're excited. The next century is going to be impossibly cool or unimaginably catastrophic. Either way, we think that there aren't enough people trying to think about what it's going to look like, and we can't afford to let it catch us off guard.

— The Editors

6

The Forecast

Modeling the End of Monkeypox

Jared Liebowich

The journalistic and public health response to the US monkeypox outbreak was noisy and contentious. What tools do we have for predicting its spread?

ILLUSTRATION BY
Kyle Ellingson

COVID-19 upended our daily lives, but for me the most anxiety-provoking part of the early pandemic was the uncertainty over how the next weeks and months would unfold: How severe was SARS-CoV-2? How contagious? Where was the next hot spot? And when would it be over? Watching the news only brought me more stress: experts couldn't seem to make up their minds about what was happening, let alone how things might play out. In trying to resolve these questions for myself, I became obsessed with forecasting.

I created an account on Good Judgment Open, a public platform where forecasters share their predictions. Trying to answer the site's COVID-19 questions helped keep me sane while dealing with the uncertainties of pandemic life. The site's founder, Philip Tetlock, is well known for identifying, evaluating, and analyzing "superforecasters" — those in the top 2% of forecasting ability. But these forecasting skills are not innate. I attribute my success in forecasting to the heuristics I've learned from thinking about hundreds of questions, as well as Tetlock's book (co-written with journalist Dan Gardner) Superforecasting. In this article, I'll demonstrate how the skills I learned forecasting COVID-19 helped me think about the spread of monkeypox in the United States — what factors went into my estimates, and where I made adjustments along the way.

Background on Monkeypox

Monkeypox is a zoonotic disease originally found in a variety of mammals, with symptoms like a milder version of smallpox. It is transmitted between humans primarily through exposure to infected bodily fluids or contaminated objects. Pockets of transmission have been recorded since at least the 1970s in Central and West Africa, but the 2022 outbreak represents the first incidence of widespread community transmission in Europe or America. The first confirmed case was reported to the WHO on May 7, but the virus was likely circulating throughout April, and perhaps even as early as March. Spain, Portugal, and the United States each confirmed cases on May 18, with many other countries to follow suit through May and June. As of this writing on September 14, cases have been confirmed in over 100 countries.

Nearly all of these cases are among men. More than 95% of cases in which data are available are among gay, bisexual, and other men who have sex with men (MSM). The preponderance of cases among MSM suggests that monkeypox is currently spreading like a sexually transmitted infection. Discussing a virus that overwhelmingly affects MSM risks reinforcing stereotypes that MSM are diseased and therefore disgusting — a form of stigma which had devastating consequences during the HIV/AIDs epidemic. Journalists and public

health experts spent the summer debating whether to de-emphasize group risk or engage in more pointed public health messaging. I won't address that debate here, but my position is that understanding how the disease spreads and who it impacts is necessary for predicting its progress and mounting an adequate public health response.

My Forecasting Process in Early August

Forecasting websites generally provide well-formed questions, but it's worth knowing what goes into generating a good question. Valuable questions tend to be ones where there are clear parameters: a discrete time period with a beginning and end, unambiguous wording, and specific sources for how a question will be resolved. For this article, I chose to forecast the following question:

How many confirmed cases of monkeypox will there be in the United States by January 1st, 2023? The question will resolve with official CDC data for confirmed cases by January 1st.

In my prediction, I looked at several critical pieces of information. First, in recognizing that the current strain of monkeypox operated more closely to an STI than other infectious diseases, I guessed that its spread would not extend beyond sexual transmission between men unless there was a significant mutation in the virus. This assumption dramatically narrows the susceptible population, putting a cap on my predicted case counts.

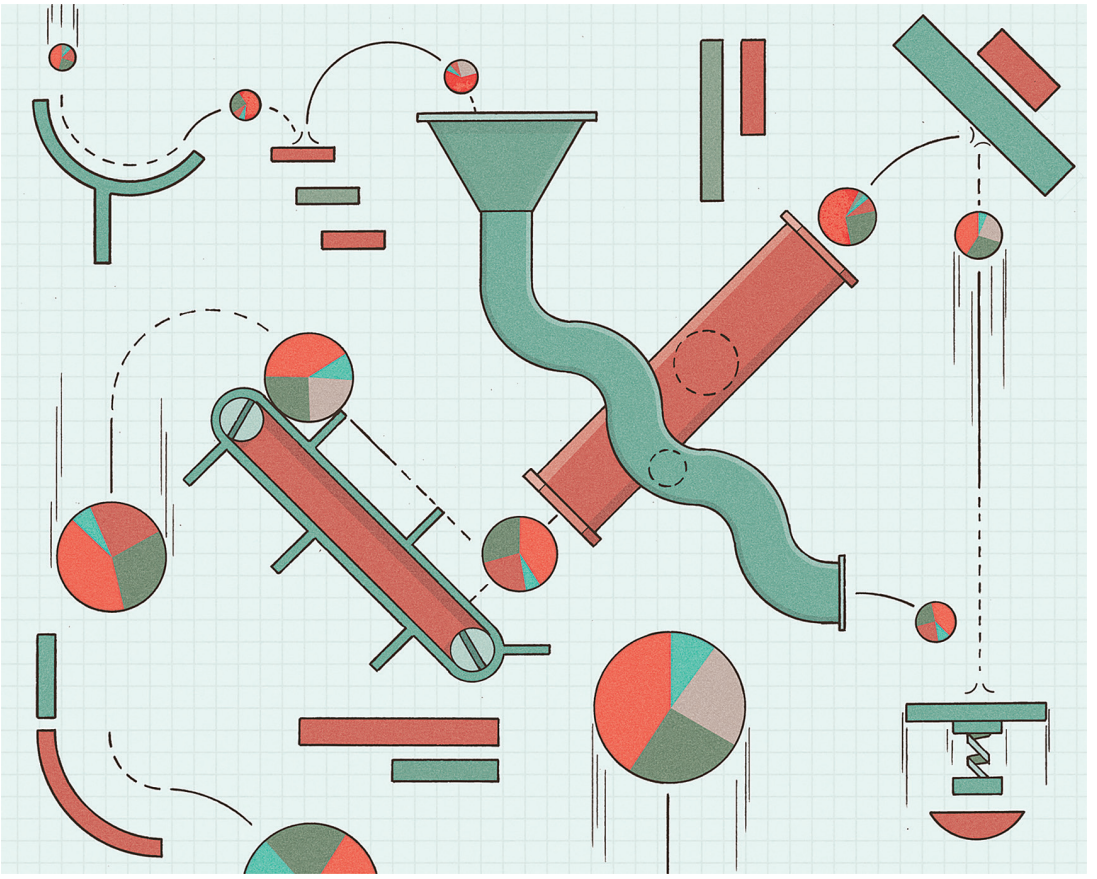
Second, I drew inspiration from Youyang Gu, a data scientist who pivoted from his day job early in the pandemic to model the spread of COVID-19. Gu's model was unique among early COVID-19 projections because it incorporated only one data input: daily deaths. From this statistic, Gu calculated parameters like reproduction number and infection mortality rate using machine learning. Gu's projections were more accurate than other models with more complicated inputs in the early stages of the pandemic. My approach was not nearly as sophisticated, and didn't apply machine learning (I used back-of-the-envelope calculations). Still, I felt confident I could make a decent forecast using only accurate data on monkeypox case rates.

Alas, such data has not been available. I made my first forecast in early August. The EU and UK were a few weeks ahead in their outbreaks, so I looked there to project trends in the United States. Here, I was hampered by inadequate data and poor reporting. Daily case counts early in the outbreak were not being released in real time. It was clear that between May and July cases had grown exponentially, but beyond that, reporting lags and high variance in day-to-day case counts made it difficult to extrapolate clear trends.

The main question I hoped to answer was when the US could expect to exit the exponential growth phase. This seemed to happen in the EU and UK in mid-August, roughly 14 weeks after the outbreak began, but sporadic reporting made it difficult to say for certain.

This left me with the American data. On June 3rd, 2022, there were 32 confirmed cases of monkeypox in the United States. By July 3rd, that had grown to 563 confirmed cases, and by August 3rd, there were 6599 confirmed cases.[1] I assumed that, given the scattered public health response

1. Due to lags in reporting, these numbers have been updated since the writing of this article on September 13. As of October 5, updated data show 66 confirmed cases on June 3rd, 1,219 cases on July 3rd, and 9,933 cases on August 3rd.



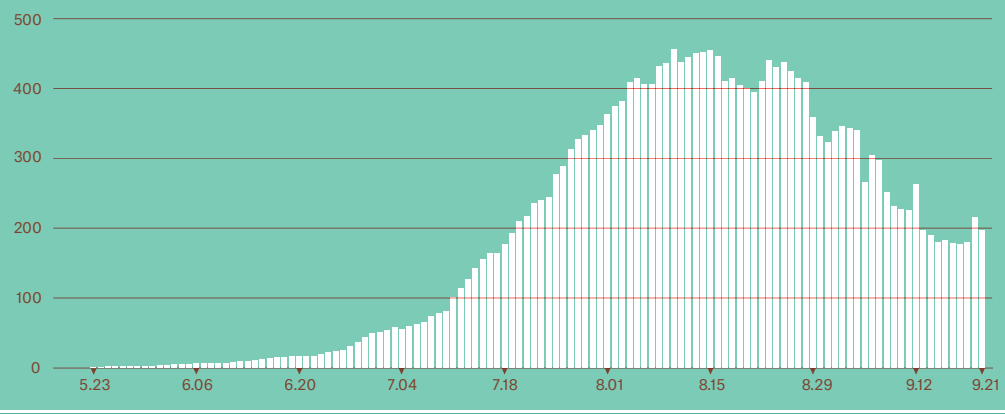
and lower awareness of the virus, June case counts were somewhat unreliable and likely under-counted, so I chose to only look at July to August spread, which saw an 11.7x increase in cases.

I assumed that this rate of growth might decrease slightly as public awareness of monkeypox increased, but would likely remain exponential. Rounding down the growth rate slightly to 10x and using the early August numbers gave me an idea of how many cases there would be by September 3 — about 66,000. I revised this downwards again based on my guess that America might exit the exponential phase in late August, leaving me with an estimate of 50,000 confirmed cases by the end of the month.

After this, I considered what would happen after America exited the exponential phase. Would new cases taper until there were basically none, or would there be a steady, linear rate of growth in total case numbers? Looking at Europe gave little indication. There, cases appeared to be potentially decreasing from exponential growth, but there was little indication of what would happen next.

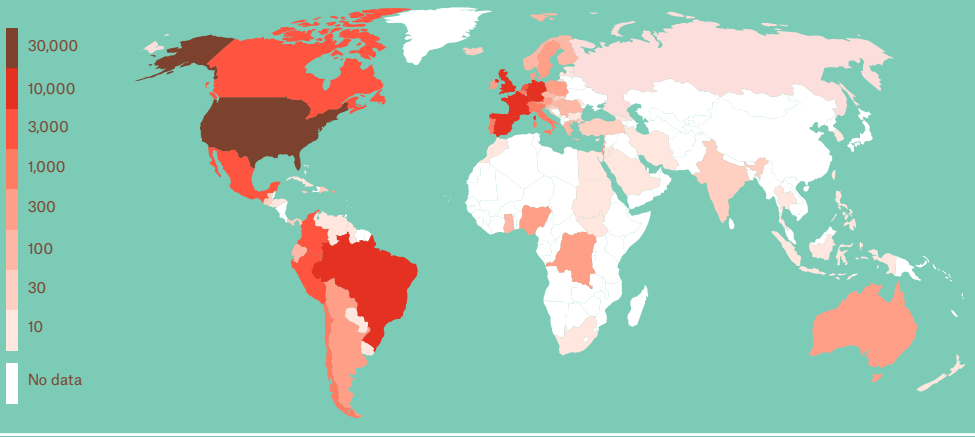
My best guess was that new cases would taper off as risky contact between MSM decreased and vaccinations increased. I also thought cases would decrease as quickly as they

Monkeypox: Daily cases reported (7-day average)



Source: Data produced by the 'Global.health' team — available at github.com/globaldothealth/monkeypox

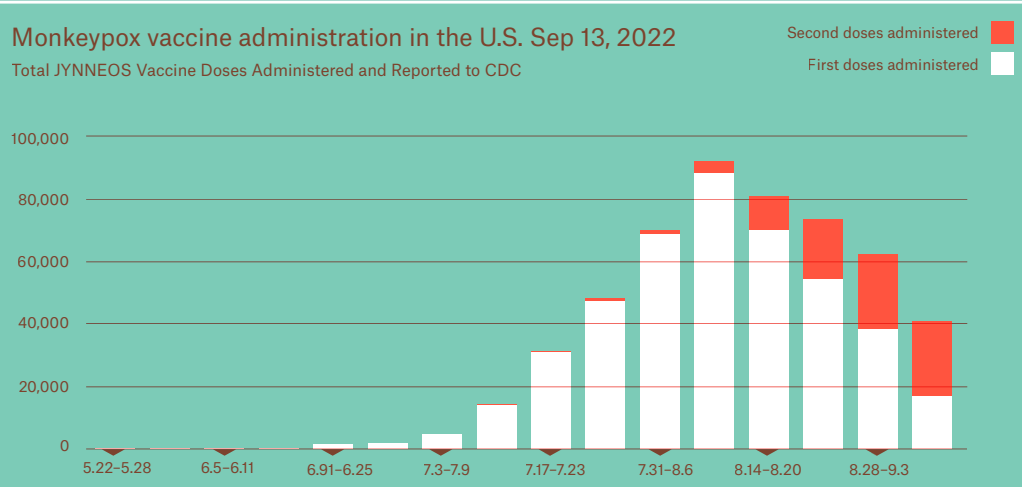
Monkeypox: Cumulative confirmed cases, Sep 13, 2022



Source: Data produced by the 'Global.health' team — available at github.com/globaldothealth/monkeypox

Monkeypox vaccine administration in the U.S. Sep 13, 2022

Total JYNNEOS Vaccine Doses Administered and Reported to CDC



Source: CDC at https://www.cdc.gov/poxvirus/monkeypox/response/2022/vaccines_data.html

increased, but I had low confidence in this assessment. Regardless, if the exponential peak occurred at 50,000 confirmed cases, I estimated that 100,000 confirmed cases by the end of the year might be likely, assuming that the exponential peak marks the halfway point in the outbreak and that cases would decrease at roughly the same rate as they had increased.

As you may already be aware, my initial forecast in August was somewhat off.

The Importance of Flexibility

An important lesson I have learned from my last few years of forecasting is to be willing to challenge my own expectations. Self-awareness and a willingness to update one's beliefs are what help to separate signal from noise, particularly for unfamiliar topics in messy news environments. It's always important to own errors. Personally, the way that COVID-19 has spread since 2020 left me primed to be fearful of pandemics breaking out. This biased me towards thinking that the monkeypox outbreak might turn into a full-blown pandemic.

Because I'd spent so much time thinking about COVID-19, I didn't consider all the ways that the two viruses might be different. And amidst all of the debate about how to talk sensitively but truthfully about the spread of monkeypox within the MSM community, I likely didn't weight a few key facts strongly enough.

First, it's very likely that a substantial proportion of the spread occurred at organized sex parties. A leading theory traces the widespread outbreak in May to several parties in Spain and Belgium. In Late July, the World Health Organization cautioned MSM to reduce their number of sexual partners, reconsider sex with new partners, and exchange contact information with new partners to enable follow-up. Many event organizers followed suit by temporarily suspending parties throughout the United States. In short, despite the debates about the muddled public health response, men who are most at risk appear to have changed their behavior in response.

Self-awareness and a willingness to update one's beliefs are what help to separate signal from noise, particularly for unfamiliar topics in messy news environments.

I underestimated how quickly this would happen. This is where comparisons to COVID-19 limited the accuracy of my first forecast. MSM as a group, particularly those who frequent sex parties, have substantially more experience in risk management around a sexually transmitted pandemic than the general population. STI testing, along with other forms of risk prevention like PrEP, are embedded in MSM communities. Moreover, activism and organizing experience gained from decades of battling HIV also played a role; MSM were quick in organizing to demand testing and

vaccination access. It's also possible that politics played a role. We're all by now well-aware of the politically and ideologically divided response to COVID-19: masks, lockdowns, vaccinations. Gay men skew heavily left compared to the general population, and political affiliation is shown to modify perceived risk (at least to COVID-19); it's possible that extends to monkeypox.

In addition, public health authorities have rapidly scaled up vaccine administration despite the initially limited supply of vaccines. In early August, the Food and Drug Administration expanded its emergency use authorization to allow health care workers to administer the vaccine intradermally, or between layers of the skin. Compared to traditional administration through subcutaneous injection — into the fat layer beneath the skin — intradermal injections use $^1/_5$ the dose. This allowed the 400,000 vials in the national stockpile to provide closer to 2 million shots. As of this writing, over 500,000 doses have been administered,[2] though demand has been steadily decreasing since mid-August.

My failure to give enough weight to these two facts highlights an important lesson in forecasting: spend time with the subject matter. There's no point trying to make predictions with only cursory knowledge of the question you're forecasting. In hindsight, these two factors were major variables; if anything, I'm surprised that my underweighting of these factors didn't lead to my forecast being more inaccurate. I think it's important not to build an overly complicated model (as this can lead to greater errors along the way), but it helps to brainstorm about any potential factors that might play a role. I am of the opinion that writing down all of your ideas, no matter how far-fetched, is valuable to the forecasting process. Choosing not to filter enables creativity — which to me is also among the most enjoyable parts of forecasting.

There's no point trying to make predictions with only cursory knowledge of the question you're forecasting.

My Adjusted Forecast

As of this writing, I assume that CDC's case trends are more accurate. They are now being updated weekly. The 7-day moving average appears to have peaked around August 10 at 461, but remained above 400 through almost all of the month. On September 9th, there were roughly 21,500 confirmed cases of monkeypox in America.[3] The level of new cases per week peaked around the first week of August at roughly 2,500 cases, then remained the same through the month. At the time of this writing on September 13th, the number of new cases each week is beginning to decline.

In my updated September forecast, I take the halfway point of the three weeks around the peak when cases flattened (August 14th) to use as a halfway point in total confirmed cases: 13,500 confirmed cases. Once again,

2. As of October 4, 873,000 doses have been administered.

3. As of October 5, the 7-day moving average reported on August 10 was 427; the peak 7-day moving average occurred on August 6 at 441. On September 9, there were 22,843 confirmed cases. The mean 7-day moving average for the month of August was just under 400.

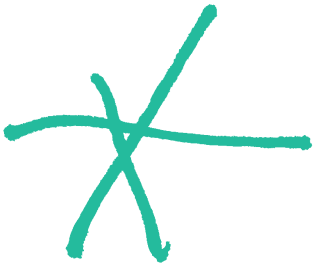
I assume that the halfway point in time in which cases have flattened serves as a midpoint in the outbreak, and that cases will decline at roughly the same rate as they have increased. Doubling this case count gives me 27,000 cases, but because I expect there will continue to be a low rate of transmission in the tail-end of the outbreak, I will increase my estimate to 30,000 confirmed cases by the end of 2022. Furthermore, I predict the outbreak will have mostly subsided by the middle of November, assuming that the 2.5 months it took for cases to reach peak transmission will be mirrored by 2.5 months for cases to decline to very low levels of transmission.

Conclusion

There are a lot of assumptions behind my forecast. Most significantly, I assume that the current monkeypox strain will not mutate into a more contagious form before the end of 2022. If a new strain arose which made transmission much easier, then cases might once again start to increase exponentially. But while forecasting COVID-19 taught me to be humble about how a virus will evolve over time, a significant monkeypox mutation seems unlikely. SARS-CoV-2 is a single-stranded RNA virus — a sloppy replicator with a high mutation rate. Monkeypox, in contrast, is a DNA virus, and does not mutate as freely. Even so, research suggests that monkeypox is mutating much more frequently than expected — 50 times in the past four years, compared to once per year for the average pox virus. So far, it's unclear if these mutations have actually made monkeypox more infectious, but public health authorities should have contingency plans in place if the strain does mutate into a more contagious or virulent form, and there should be in-depth monitoring of the viral genetic sequence for confirmed cases, as we've done with COVID-19.

The monkeypox outbreak may have been preventable. Tests and vaccines were on hand before the outbreak reached US soil and local authorities were initially slow to respond to demand. Understaffed state and local health departments and a fragmented reporting system has hampered coordination between those authorities and the CDC — and as I've said, impeded early attempts to forecast the outbreak.

Despite all this, the monkeypox outbreak has been less severe than I initially predicted. Modeling COVID-19 was difficult for many reasons. The virus continues to evolve. Vaccine efficacy has ebbed with new variants. Our response to the pandemic — human behavior — was perhaps the most difficult variable of all to model. The biggest uncertainty in my model was when the monkeypox outbreak would exit exponential growth, and it happened much earlier than I expected. I attribute that relatively fast decline in cases to the organizing and advocacy of the MSM community. After two and a half years of COVID-19, that sort of unified action was difficult to predict.



14

Review

What We Owe the Future

Kelsey Piper

William MacAskill's latest book presents itself as an introduction to the burgeoning longtermist movement. But his views are eccentric— even within the movement he founded.

Basic Books
August 16, 2022

I have some questions about *What We Owe the Future*.

In this, I'm in good company. The book debuted with a spot on the New York Times bestseller list; a profile of MacAskill in the New Yorker; and reviews in the New York Times, the Wall Street Journal, the Guardian, Salon, the Boston Review and many more. MacAskill was interviewed on, and I exaggerate only slightly, all the podcasts. Ads for the book were everywhere. The contrast with the successful, but modest, launch of MacAskill's previous book, *Doing Good Better*, is a window into something bigger: The effective altruism movement MacAskill founded is resourceful and well-connected these days, and can get the key things it has to say before millions of eyes.

MacAskill's book has met with a broadly positive reception from the general public. Interestingly, though, the reaction from the effective altruism movement has been mixed. Effective altruists who started out skeptical of MacAskill's longtermism are still skeptical. Perhaps more surprisingly, effective altruists who share his worldview still objected to many of the book's details, pointing out that its perspective on what priorities are implied by longtermism is out of line with what most longtermists — other than MacAskill — are actually doing.

What is the longtermist worldview? First — that humanity's potential future is vast beyond comprehension, that trillions of lives may lie ahead of us, and that we should try to secure and shape that future if possible.

Here there's little disagreement among effective altruists. The catch is the qualifier: "if possible." When I talk to people working on cash transfers or clean water or accelerating vaccine timelines, their reason for prioritizing those projects over long-term-future ones is approximately never "because future people aren't of moral importance"; it's usually "because I don't think we can predictably affect the lives of future people in the desired direction."

As it happens, I think we can — but not through the pathways outlined in *What We Owe the Future*. MacAskill is a philosopher, and *What We Owe the Future* is a philosopher's book, satisfied at times with an existence proof: Look, he likes to say, abolitionists affected the next century, perhaps the next several centuries; Confucianism won out over other Chinese philosophies and then held on for thousands of years; empires collapse and societies stagnate. You can try to change the morals and values of the society around you, and under some circumstances your changes will have ripple effects into the distant future.

Under *some* circumstances. But under *which* circumstances? How do you know if you're in them? What share of people who tried to affect the long-term future succeeded, and what share failed? How many others successfully founded institutions that outlived them — but which developed values that had little to do with their own?

The first and most fundamental lesson of effective altruism is that charity is hard. Clever plans conceived by brilliant researchers often don't actually improve the world. Well-tested programs with large effect sizes in small, randomized, controlled trials often don't work at scale, or even in the next village over. Some interventions manage to backfire and leave recipients worse off than before — MacAskill's favored example of this, described in his book "Doing Good Better," is PlayPumps, an expensive and ill-conceived plan to replace standard water pumps with pumps children could operate while playing (they break down easily, and extracting useful manual labor from children turns out not to work very well).

Not only is it fiendishly difficult to do something that works, it's often even harder to tell when it has. EA charity evaluator GiveWell has been trying for more than a decade now to figure out how cost-effective it is to distribute deworming medication to children, and their error bars have only narrowed a little from when they first began. Are graduation programs better than cash? Depends how you measure it.

Most well-intentioned, well-conceived plans falter on contact with reality. Every simple problem splinters, on closer examination, into dozens of sub-problems with their own complexities. It has taken exhaustive trial and error and volumes of empirical research to establish even the most basic things about what works and what doesn't to improve peoples' lives.

These questions are not unanswerable. Through the heroic work

of teams of researchers, many of them have been answered — not with perfect accuracy, but with enough confidence to direct further research and justify further investment. The point isn't that everything is unknowable; the point is just that knowing things is hard.

That is, ultimately, the simple yet damning response to *What We Owe the Future*: It does not actually convince me that it has any proposals that matter on the cosmic time scales that it speaks of. This is the fundamental challenge which longtermists must rise to, and which *What We Owe the Future* has to answer.

Most well-intentioned, well-conceived plans falter on contact with reality. Every simple problem splinters, on closer examination, into dozens of sub-problems with their own complexities.

And viewed through that lens, it's a somewhat unsatisfying book. There are a lot of people who've attempted to change the world — through conquest, through science, through politics — and plenty who look to have succeeded on the scale of at least a few hundred years. But almost none of the weight of MacAskill's arguments applies to changes on the scale of a few hundred years; they largely rest on the possibility that our actions have impacts on the scale of millenia.

This is most evident where the book addresses the possibility of technological stagnation, where the position it takes is unique: MacAskill argues that because we will eventually run up against hard technological limits to the size of our economy, speeding up economic growth might not matter much — having large effects over how the next few thousand years go, but minimal effects on where we ultimately end up.

Of course, if causing a thousand years of technological stagnation isn't a significant act, little is. MacAskill's section on moral change argues that slavery could have persisted into the present day in the absence of a dedicated and well-organized abolitionist movement. The end of mass slavery, he argues, wasn't overdetermined as a result of economic changes; it was largely the product of a specific political campaign to throw the might of the British Navy behind ending the slave trade. If that group hadn't acted, slavery could have endured much longer — at least until someone acted with the conviction and determination of the abolitionists in a similarly hospitable political climate. It's not hard to imagine that could have taken decades or even centuries.

It's impossible to read the full account and not feel awed by the determination and conviction of

the early abolitionists, or gratitude for the much-better-than-it-could-be world they left us.

But on MacAskill's own terms, it's hard to claim abolition as a longtermist achievement — an astonishing humanitarian triumph of principled political organizing, yes, but one which mostly justifies itself through the benefits to already-alive enslaved people and their children and grandchildren, not through the benefits to future human civilization.

Many of the same questions come up in the examination of the founding of the United States, which I'm happy to call one of the biggest longtermist success stories in history (though MacAskill doesn't go quite that far): The Founding Fathers envisioned, and mostly created, a country with distinctive political and social commitments that have eventually made it a superpower. But there are still a dozen questions: How many governments did people found with similar intent that didn't work out? (How many in France alone in the same time period?) How good does America actually look today, from the perspective of the founders? Is that about as much influence on our descendants three hundred years hence as we can really hope for, absent unusual technological situations? What applicability does this example even have to people who aren't in the position of starting a revolution and founding a new government? Will the influence of the U.S. really extend eternally into the future?

There is a way of making sense of this set of commitments, which the book briefly gestures at (though I think without a lot of background familiarity with longtermism, the connection between these arguments would be nearly impossible to parse). If it so happens that modern-day humanity builds superintelligent machines, and these machines use our current values to steer civilization, then anything that affects contemporary values and governance will end up having outsized long-term impacts. That's the "values lock-in" argument, and it's the main reason to think that if we change peoples' priorities and moral commitments today it could affect the distant future — which is to say that if we don't buy the values lock-in argument, there's little case for trying to change peoples' values on longtermist grounds.

You might expect people to depart from MacAskill at the claim that superintelligent AI will transform the world in our lifetimes, but, in fact, he is in good company. Effective altruists have worried about emerging technologies that could make it easier to wipe out all of human civilization since the movement was founded, and AI is a major research focus and priority.

But most EAs working on AI disagree with MacAskill about precisely where the challenge lies, a fairly technical disagreement with major implications for what longtermists should do today. MacAskill thinks we're fairly unlikely to just lose control of the future to AI systems that have no reason to do what humans ask; in a footnote he rates that likelihood at around 3% this century. (Most

longtermists are substantially more pessimistic than that.)

The more typical longtermist perspective is something like this: Broadly, current methods of training AI systems give them goals that we didn't directly program in, don't understand, can't evaluate and that produce behavior we don't want. As the systems get more powerful, the fact that we have no way to directly determine their goals (or even understand what they are) is going to go from a major inconvenience to a potentially catastrophic handicap.

For this reason, most longtermists working in AI safety are worried about scenarios where humans fail to impart the goals they want to the systems they create. But MacAskill thinks it's substantially more likely that we'll end up in a situation where we know how to set AI goals, and set them based on parochial 21st century values — which makes it utterly crucial that we improve our values so that the future we build upon them isn't dystopian.

Which of these failure modes is more likely has major implications for which approaches to securing the future are most promising. If you think humanity is likely to fail catastrophically at designing AI systems that have goals we can understand and influence — and therefore likely to unleash AIs whose values bear little resemblance to our own — then improving present-day human values isn't a longtermist priority: The important thing is making sure humanity gets a future at all. If you think large-scale transformative effects from AI aren't likely to happen then as a longtermist you'd probably focus on other sources of existential risk, rather than on any of the values changes or AI-related risks MacAskill highlights.

Effective altruists have worried about emerging technologies that could make it easier to wipe out all of human civilization since the movement was founded, and AI is a major research focus and priority. But most EAs working on AI disagree with MacAskill about precisely where the challenge lies.

Either way, "What should longtermists do?" is a deeply technical question that depends on assessments not just of whether AI is going to pose an unprecedented threat, but of exactly how it's going to do that — a question where MacAskill happens to disagree with most others focused on risks from AI.

Perhaps this is why the advice MacAskill gives for how to put longtermist principles into action feels disappointingly scant. He recommends voting, being a "moral weirdo" and pursuing careers in effective altruism, but without a clear unifying thread for how those avenues produce the kind of rare and distinctive long-term changes in the world that are inspiringly profiled in the first half of the book.

Many of the critiques of *What We Owe the Future* have gestured at this complaint, sometimes indirectly. "There's more than a whiff of selective rigor here," Christine Emba complained in the Washington Post, and she's right: While *What We Owe the Future* is almost fanatically cited and fact-checked, with appendices that rival the book in length (that's a compliment), the treatment of ways to affect the long-term future often seems satisfied with the fact that doing so is possible rather than estimating the actual odds of success.

MacAskill doesn't actually make the argument that, Pascal's wager style, we should be pursuing minute possibilities of influencing the long-term future because even a one-in-a-billion chance of affecting trillions of future people is so important. I happen to know he doesn't believe that. But it feels to me like an obvious consequence of where the book chooses to focus. It does a remarkably compelling job of introducing and driving home the sheer magnitude and potential of humanity's future, and then offers a light survey of things that might influence it, without many strong arguments about which are the most important.

Here, again, most EAs are doing something different: They'd be happy to tell you what the most important longtermist work is, and they generally think it's preventing extinction. In biosecurity and AI, rapidly advancing technology will make it cheaper and easier to cause mass death on an unprecedented scale (and make it likelier to happen by accident); that is where the bulk of EA money and attention not directed at present-day causes is directed. For many of these people, their life's passion is making sure a specific new technology doesn't kill us all in the next 100 years; they are, in fact, often doing this out of their deep concern for the trillions of people who might live in a flourishing human future, but they're only trying to influence *events* they expect to occur quite soon.

The stakes are as high as MacAskill says — but when you start trying to figure out what to do about it, you end up face-to-face with problems that are deeply unclear and solutions that are deeply technical.

MacAskill tackles risks to human civilization in Chapter 5, titled Extinction, though I don't think he makes what I consider the strongest argument for it: Trying to prevent extinction in the next century possesses at least some of the concreteness that values change lacks (especially if you don't expect values lock-in in the near future and so have to worry about how long your values differences endure).

You can try to invent specific things that make extinction less likely, like (in the case of pandemic preparedness) better personal protective equipment and wastewater surveillance. You can identify things that make extinction more likely, such as nuclear proliferation, and combat them. These are still thorny problems that reach across domains and in some respects confuse even the full-time experts who study them, but there are achievable near-term technical goals, and longtermists have some genuine accomplishments to point to in achieving them.

In the short term, persuading people to adopt your values is also concrete and doable. Effective altruists do a lot of it, from the campaign against cruelty to animals on factory farms to the efforts to convince people to give more effectively. The hard part is determining whether those changes durably improve the long-term future — and it seems very hard indeed to me, likely because my near-term future predictions differ from MacAskill's.

That's how I end up agreeing 99% with a worldview but feeling profoundly mixed about the book that lays it out. The stakes are as high as MacAskill says — but when you start trying to figure out what to do about it, you end up face-to-face with problems that are deeply unclear and solutions that are deeply technical. Is MacAskill right that we are likely to build AI systems that have human-set goals but the wrong human-set goals, or am I right that we're likelier to fail by not knowing how to set their goals at all?

I think we're in a dangerous world, one with perils ahead for which we're not at all prepared, one where we're likely to make an irrecoverable mistake and all die. Most of the obligation I feel toward the future is an obligation to not screw up so badly that it never exists. Most longtermists are *scared*, and the absence of that sentiment from *What We Owe the Future* feels glaring.

If we grant MacAskill's premise that values change matters, though, the value I would want to impart is this one: an appetite for these details, however tedious they may seem. The stakes are high. The problems we're trying to think about are without precedent and deeply weird. The problem is that *What We Owe the Future* doesn't quite feel like it reflects a style of thought that'll get us to the bottom of them.

22

Interview

Making Sense of Moral Change

Christopher Leslie Brown

A conversation about abolitionism, moral progress, and the pitfalls of historical counterfactuals.

Asterisk: *Your book* Moral Capital *is about why the movement to abolish the slave trade in Britain happened in the late 1780s and not earlier. Would you mind briefly walking through the thrust of that argument?*

Christopher: While it's not easy to boil down the entire book, essentially, there's a group of people who gather in the late 1780s and commit themselves to convincing British authorities to abolish Britain's slave trade. The book explains how that group came together, who they were and why they chose that particular issue. The broad answer is that the circumstances of the American Revolution and its aftermath created an environment with new political, moral and cultural values that did not exist before. I don't argue that the American Revolution caused the antislavery movement, but that it created the conditions that made the movement possible.

A: *The American Revolution politicizes it.*

C: Exactly right.

A: *In the 1770s, America and Britain both placed the blame for slavery on each other. American patriots blamed British slave merchants. The British blamed American plantation owners. Yet despite that, very little changed in the short term.*

C: What that indicates to me is that there was recognition on both sides that slavery was a problem. You can't put blame on somebody for something if you don't think there's something blameworthy happening. So while there's a collective acknowledgment on both sides that slavery is morally and ethically questionable, that gets weaponized. There's a schoolyard quality to it: It's not my fault, it's your fault. But I think we can recognize those arguments were, in one form or another, guilty consciences at work.

A: *That makes me think about a fact I found surprising. You have many examples of people who, when they are exposed to information about the Atlantic slave trade, responded in a way that seems normal to our modern sensibilities. They're shocked, they're horrified, they see slavery as evil. But then this doesn't translate to any action on their part. There's a disconnect.*

C: I think one of the main contributions of the book is to explore the gap between values, ideals and principles and feeling compelled to act on them. That gap is an idea I'm very committed to. It came out of my reflections on what I regard as everyday ethical and moral experience: We may hold certain ideas and values, but it can be very difficult to align our lives with our values. People often only do it selectively or partially.

Ideals are just that — ideals. They're not about the lives we live every day. We see this in our own lives all the time. We're exposed to a piece of

negative or troubling news and we think "Isn't that terrible?" But then we go about our days. We go to sleep at night having chosen, for one reason or another, not to act.

In other scholarship on slavery, authors have tended to assume that once ideas and values are established, you can make sense of why people acted. *Moral Capital* is about showing just how wide the space is between holding a view and acting on it. My own view is that it's actually unusual when people are mobilized around their notion of what is ethical. I think the more common experience is to find ways to justify things that are unpleasant or uncomfortable because it's hard to know how to act on them.

A: *Antislavery sentiment is widespread by the 1780s — and perhaps even earlier. Can you describe that history?*

C: It doesn't require a great deal of moral or ethical insight to see that treating a human being as a thing, not a person, is wrong. Just about every sort of spiritual, ethical or moral system has described slavery in one form or another as "against nature."

But historically, people also thought of slavery as a product of civilization and therefore found ways to justify it. They created ways to explain, for example, why specific groups of people should be enslaved, or why a particular system of slavery is justified, even a form of progress over a previous system. Anybody reflecting on this (and this is true in any of the times where slavery has been predominant) can see that slavery is wrong, but they would also see that that's how the world works.

I've always thought that the best analogy is how we regard eating meat. With modern science and the ability to manage food systems, we can nourish ourselves without killing animals. We also know that eating animals is unethical. We're doing something cruel and unnecessary. But we do it because that's just what we do. It wouldn't be surprising to me if 30, 50 or 120 years from now, people look back on this time period and ask, "What was wrong with these people? They must have been like moral infants. They must not have realized this is a cruel way to treat animals." But we know it perfectly well. It's just what we do. I see slavery similarly: cruelty and brutality backed by all sorts of justifications that are built into the age.

A: *I don't eat pork or chicken for animal welfare reasons. And the line of thinking that convinced me to actually do this was that if I was alive in the 18th century, I'd want to be the kind of person who's boycotting sugar.*

C: Right. A lot of this functions in part because the cruelty, the brutality — you might even say the inhumanity — is invisible to us. Meat simply shows up in a store or on a plate. We're not confronted with its creation. That's one reason why the attack on the slave trade focused on making visible what everybody knew but didn't want to think about.

A: *As we know, at some point, attitudes do begin to shift. The 18th century saw the emergence of new ways of thinking about morality as a result of, for example, the Enlightenment, the rise of humanism, a greater attention to sentiment and even evangelical Christianity. How did these impact thinking about morality? And do you think these developments were genuinely new?*

C: I do. When I wrote the book, I was more skeptical about this line of thought than I am now. I think that there is a notable impact of those emerging worldviews that deepened the antislavery biases that already existed.

First, there's a push against certain legal explanations of slavery. Specific legalistic ways of thinking about the grounding for slavery come under withering attack in the second half of the 18th century. That makes a difference. Second, there's a new value assigned to emotional response — a newly positive association with deeply intense feeling, as opposed to regarding emotions as something to control or contain. Those cultural developments are significant.

A: *Can you expand a little bit on this new relationship to legalism?*

C: One justification for slavery in the Americas was that those who were enslaved had been captured in a "just" war and that slavery was therefore a "just" substitute for death. That rather than being killed on the battlefield, lives were turned over to their captors. Essentially, that the enslaved had been spared. This is a notion which is in Roman law historically and was prominent especially throughout the Spanish American context.

One of the things that happens in the middle of the 18th century is that this idea of slavery as a substitute for death or captivity grows to be regarded as ridiculous. It becomes treated as a fiction — and a convenient fiction at that. That's a big piece of the change.

A: *This change that's happening in the late 18th century seems to be coming from every direction. There are people whose antislavery sentiment is an expression of enlightened, humanistic thought about human society. There are people for whom it is an expression of their desire to either prop up or attack the British Empire. And still for others it's because they're deeply Christian.*

C: One reason why antislavery becomes a movement is because it has value for very different groups of people who, in one way or another, see it as a vehicle for advancing broader issues that concern them.

A: *What do you see as the common threads between all of these different ideological and cultural factions?*

C: First is the fundamental discomfort with the enslavement of other people. There's an important cultural difference between Europe and the

With the exception of Anthony Benezet, I don't think there is anyone who in their opposition to slavery, did not have some other purpose in mind that mattered as much or even more to them. That doesn't make them insincere. It just means that when people are doing things, they're usually bringing all of themselves to it.

Americas. European settlers in America are conscious about developing societies that are different from European societies. That divergence ends up having political implications for both sides, especially once the relations between become so fraught.

The other thing I think is worth saying (although this is more contextual than it is circumstantial) is that the slave trade just explodes in the middle decades of the 18th century. I'm not a big believer of threshold explanations. I don't think you can say that at a certain level it necessarily becomes a moral and political issue. But there's no question that the movement against the slave trade emerges at the trade's height, not at a point where it was declining in some form or fashion or was becoming less useful strategically. It emerges in the context of a flourishing commerce.

A: *At the core of the movement to abolish the slave trade are the Quakers. They provide a lot of the motive force and a lot of the organizational force. What makes the Quakers so special?*

C: One thing that makes the Quakers special is their independence. The Quakers develop a set of practices and values that are quite distinct from their neighbors. They take pleasure in this distinctiveness and they enforce it. A key feature of Quaker distinctiveness is that they never develop anything like a theological orthodoxy. That's different from most other Christians and other religions in the Western world. For the Quakers, there's no one set of beliefs that one has to subscribe to.

Instead, Quakers hold a set of values — like peace or the universal access to the Inner Light.[1] As a result, it's easier for changes to emerge from within Quaker life than it is in other systems where hierarchies serve

1. Quakers believe that silence and stillness enable people to access what they refer to as Inner Light, which shows them truth about their lives and how they should act to align themselves with God's will.

as gatekeepers to what can be said. There are always dissenters in every tradition, but Quakers make more room for dissenters than others do. In addition, the Quakers, from their founding, have a real ambivalence about wealth and hold a hostility to violence and to war. Slavery produces an enormous amount of wealth. And slavery can't operate without the regular practice of violence.

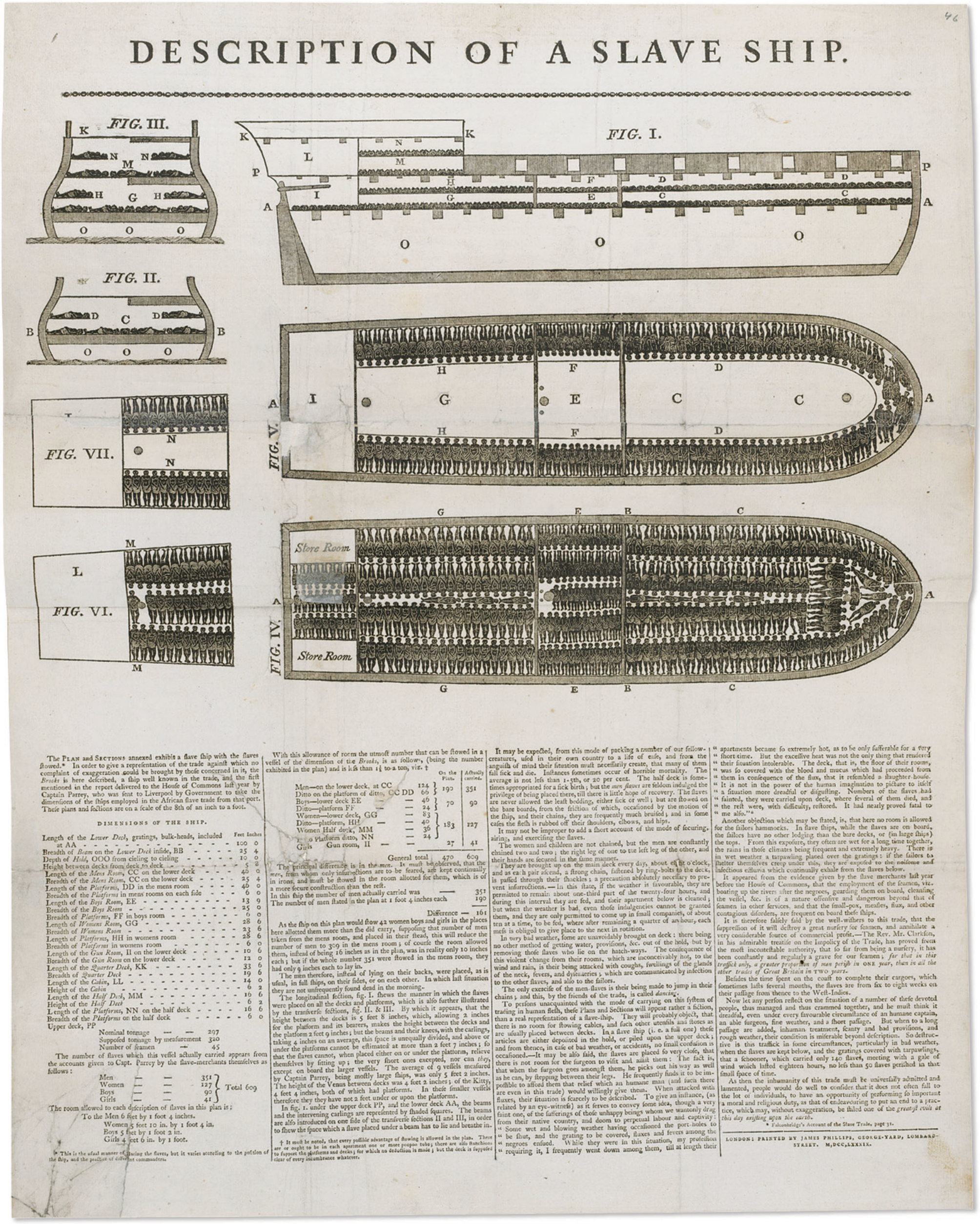
But, as most people know, Quakers were slaveholders. Quakers owned slaves from the very moment they arrived in the Americas. But when purists within the Society of Friends began to take the view that slaveholding is un-Quaker, they were able to exercise influence, which is unusual in other religious denominations. It became a kind of purity test for the Quakers. Initially, what Quakers were concerned with was reinforcing the specifics of their religious witness. Withdrawing from the slave system was a way of reinforcing what it meant to be a Quaker. Then they decided, for reasons that have to do with the political culture around the American Revolution, to carry that witness out into the world. But initially, Quakers were trying to reinforce a collective identity.

A: *And this goes back to another major theme of the book: that there is a false dichotomy between sincere activism and self-interested activism. Abolitionists were quite sincerely horrified by slavery and motivated to end it, but their fight for abolition was not entirely altruistic. They were also benefiting, socially, from the fight.*

C: This is one of my deepest convictions. I don't think it gets enough attention. There's been a way of writing about abolition, at least until *Moral Capital*, as if people woke up in the morning as abolitionists, went about their day as abolitionists, and fell asleep as abolitionists — that they maintained a consistent antislavery identity. But that's not how anybody lives their life. People had to come to the issue for one reason or another, and it had to relate to the other things that mattered to them. The issue of slavery was connected to the other things that abolitionists were interested in. With the exception of Anthony Benezet,[2] I don't think there is anyone who in their opposition to slavery, did not have some other purpose in mind that mattered as much or even more to them.

That doesn't make them insincere. It just means that when people are doing things, they're usually bringing all of themselves to it. They're bringing their social networks, their personal experience, their hang ups, their preoccupations. The search for the perfectly selfless person really misses what is in fact a complex set of motivations that move people to

2. Anthony Benezet (1713-1784) was a American Quaker abolitionist who founded the Society for the Relief of Free Negroes Unlawfully Held in Bondage, one of the world's first antislavery organizations. "Human equality to Benezet was an ontological fact rather than a philosophical doctrine or maxim. He faithfully exhibited in practice the social and moral obligations that followed from these values." *Moral Capital*, 397.



This diagram of the 'Brookes' slave ship was created in 1787 and widely copied and distributed by abolitionists. The image illustrates how enslaved Africans were transported to the Americas and depicts a slave ship loaded to its full capacity — 454 people in the hold. Thomas Clarkson commented that the "print seemed to make an instantaneous impression of horror upon all who saw it, and was therefore instrumental, in consequence of the wide circulation given it, in serving the cause of the injured Africans."

act. Everyone operates between these poles of selflessness and self-interest. It's in that space that motivations emerge and decisions get made.

A: *This is actually something I think about a fair amount in the context of effective altruism. Like any other community of people who are trying to do good in the world, you can get more respect and status if you're living out the community's values.*

C: Yes. I think Thomas Clarkson[3] is the most important person in England for the antislavery movement from beginning to end. When the movement is established, he's 23 or 24 years old and he thinks this is how he's going to make his name for himself. This is how he's going to distinguish himself.

If you decide to lead a movement, you have to think of yourself as someone who can lead a movement. You have to have grandiosity and a notion of having a capacity for unusual efficacy and leadership if you're going to do something like that. I don't see how you can be a major actor in the world if you're not on an ego trip.

A: *There's a great line you have about Clarkson, that he's obsessed with finding proof of his own genius.*[4] *I know a lot of people like that, and many of them have done an enormous amount of concrete good.*

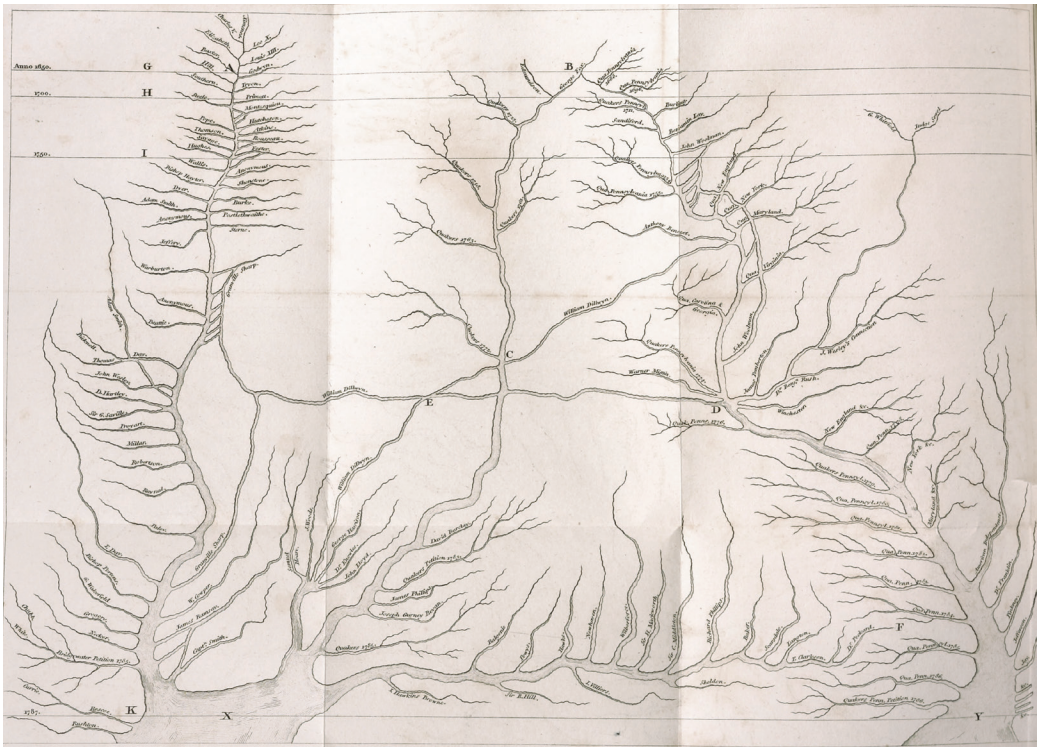
C: Exactly. Part of what I'm trying to do is to recognize — even destigmatize — that sort of self-absorption because those are the people who make unusual impacts in the world. They're often not pleasant to be around, but you have to recognize the balance of virtues and vices. Even thinking about it on a local level, imagine someone who says, "I can run for mayor and change the city," or, "I can take over as principal and make it the best high school in the district." To say that is to say that you have something special that other people don't have or won't act on. Those two ideas can't be separated. That sort of self-importance often rubs us the wrong way, but one of the purposes of the book is to make clear that heroism is grounded in some of the mucky stuff of being human.

A: *And if you are an enslaved person, I cannot imagine that you care what motivates Thomas Clarkson.*

C: I think that's right. Although there's one thing I think often gets a little bit overlooked, and I didn't make enough of this in the book: None of

3. Thomas Clarkson (1760-1846) was an early organizer against the slave trade in the British Empire. He helped found the Society for Effecting the Abolition of the African Slave Trade and was instrumental in the passage of the Slave Trade Act of 1807, which ended British trade in slaves.

4. *Moral Capital*, 439.



This image, from Thomas Clarkson's 1808 book *The History of the Rise, Progress, and Accomplishment of the Abolition of the African Slave-Trade by the British Parliament*, illustrates the intellectual streams which flowed together into the abolitionist movement. The top right of the map shows Quaker activists, including Anthony Benezet and Benjamin Lay, while the top left shows Enlightenment thinkers such as Montesquieu and Rousseau. Clarkson himself can be seen at the bottom right.

these people took an interest in or really asked enslaved people what they wanted for themselves.

A: *Even Benezet?*

C: He's at the outer edge of it. It's a good point. Let's just put it this way — he didn't write about it. He did not try to ventriloquize and say, "Here's what enslaved people are saying." He did not write "Here's what I was told." He was an advocate with a constituency, but speaking and pushing for what *he* thought. None of the abolitionists actually thought that slaves should be not only free, but fully equal and able to do whatever they wanted. Nearly all of the perspectives on abolition still expected enslaved persons to work — just without slavery. Antislavery did not mean pro-Black or pro-African, and certainly not anti-racist. You can oppose slavery and still believe in racial inferiority, as just about everyone did in some form or fashion on the European side.

The image of a kneeling slave asking "Am I Not a Man and a Brother?" was designed by Josiah Wedgewood's pottery firm as a seal for the Society for the Abolition of the Slave Trade in 1787, and soon became an international symbol of the abolitionist movement. It appeared in books and pamphlets, on jewelry, snuffboxes, and ceramics, and in a series of medallions made by Wedgewood made for members of the society.

A: *Changing the subject — The Society for Effecting the Abolition of the African Slave Trade is founded in 1787, and the slave trade is abolished in 1807. And then it's another 10 years before activists start even trying to get emancipation for existing slaves. Why is there that gap?*

C: There was an expectation that once the slave trade was abolished, it would lead to a steady decline of slavery itself. That didn't happen. Initially there was a fear that British slaveholders in the Caribbean were smuggling in enslaved men and women by other means. So there developed a proposal to create a national census or registration, so authorities could make sure that the slave population wasn't increasing.

Then you had a phase of amelioration: that slavery needs to be made kinder and more gentler. Then when the Caribbean slaveholders didn't cooperate with that idea in any meaningful way, you start getting pushes for first gradual emancipation, then immediate emancipation.

Library of Congress

Some thought that abolishing the Atlantic slave trade would be a first step toward emancipation, but that was not how the vast majority of the abolitionists talked about it. Many of them weren't thinking about it that way. What they thought it would do is start a virtuous cycle that would lead to the changes they hoped for.

A: *I wanted to get to the topic of historical contingency. This is something you talk about a lot, and of course it's very relevant for anyone thinking about how — or if — we can influence the long-term future. What do you think happens with the abolitionist movement if there is no American Revolution, or if the American Revolution gets pushed back?*

C: I don't think abolition would have been politicized in the same way without the American Revolution. That's really the main issue. There's a certain kind of conspiracy of silence around slavery in the British Empire because it benefits everyone except for enslaved people. It's only when a divide forms within the political elite over what's happening in America that it becomes useful to project and assign blame somewhere else. I think it's contingent in that sense.

It's possible there could have been a push for gradual reforms, but the sensitivity of slaveholders in North America and the Caribbean would likely have been too high. The idea that anyone in London could have any say over slaveholding would have been too threatening. Given the way that North Americans responded to the Tea Act — which, it's worth remembering, actually lowered the duties on tea — you can only imagine what their reaction would have been if the British government had started putting rules down about how Americans could treat enslaved people. The movement was premised on a division.

Having said that, I do think it's possible that some of the Northern states might have acted on their own. Massachusetts, in particular, very well could have decided that they were going to outlaw slaveholding.

A: *And then by 1850s, '60s, '70s, places in Spanish America and the rest of the world are beginning to outlaw slavery — how does that fit into this story?*

C: Today, no one argues we should restore slavery. It's illegal everywhere on the planet, even though it operates in the shadows. There's a consensus by the 21st century that slavery is beyond the pale. But I don't think that was inevitable. I'm not saying that it was unlikely or highly contingent, but I think it's incorrect to think about abolition in the Americas or today's global consensus against slavery as if it's part of the natural process of modernization. Marx, for instance, treated slavery as a stage in economic development that advanced societies would grow out of. I don't think any of that is true.

I think slavery could have lasted well into the 20th century, despite the advantages of industrialization and mechanization. I don't think abolition would have occurred without the pressure and power of the British

government's backing. No countries abolished slavery independent of the example or the force of the British Empire. Britain in the 19th century was the most powerful and influential nation in the world.

A: *So these late 18th century British activists really do have this globally outsize impact.*

C: The movement became an international force after 1815, following Napoleon's defeat and Britain's emergence as a global superpower. Britain abolished not only their own slave trade, but the global slave trade. Absent that, there's likely no domestic movement for the abolition of any of the other slave trades — for example in Portugal, France or the Netherlands. Those governments are forced to do it because they're faced with British diplomatic or military pressure. That's why the British case is so important.

A: *This implies then that there is also the potential for changes in the US to be overly impactful.*

C: We obviously have the largest global military footprint. Our economy has an outsize impact. In one way or another, every country has to maintain some relationship with the United States and American norms. It was the same with Britain in the 19th century. In the same way that the United States "exports democracy," Britain exported antislavery — even if that was somewhat contradictory to the reality on the ground.

A: *Is there anything else that you think is relevant that I should be asking you that you want to bring up?*

C: Counterfactuals are tricky things. I know effective altruism is in part a forward-looking enterprise. On the one hand, I do believe that historians have a special contribution to make in thinking about how the world that we inhabit might evolve. On the other hand, I also really believe our gifts are retrospective, not prospective. One of the purposes of retrospection is to attend to the individual and the specific.

I believe, philosophically, in the butterfly effect. Small differences can make large changes. I really do believe that. I think if Thomas Clarkson had died of smallpox when he was six, we might still have had an antislavery movement in Britain, but it would have evolved differently. It would look different.

I do believe that the peculiarities of individuals make a massive difference. But I don't think you can say how things would have played out in the counterfactual. There's no way to know.

34

Interview

How to Prevent the Next Pandemic

Kevin Esvelt

Modern technology makes bioterrorism seem increasingly likely. If we can get our act together, there are smart ways to prevent it.

ILLUSTRATIONS BY
Claire Merchlinsky

Asterisk: *Can you tell me a little bit about yourself and your lab and the work that you do there?*

Kevin: I am an associate professor at the MIT Media Lab, which is a place for people whose work does not fit in any single discipline. At my lab, called the Sculpting Evolution Group, we are interested in advancing biotech safely. We study the evolution of molecular systems over time and ways of applying selective pressure to make them do what we want and keep doing what we want.

I also have a bit of a security mindset. In cybersecurity there's a saying: no system vulnerable to accidents is helpless against deliberate attack. Wherever it came from, SARS-CoV-2 was an accident. It was either a natural or accidental release, but it was not deliberate, because anything deliberate would be more severe. That suggests that if and when we learn how to build harmful things with pandemic-class capabilities, we're going to be in trouble. Lots of people are going to be able to cause pandemic-class events, and the rest of us are not going to be able to do much to defend against them.

A: *Why are you so concerned about this possibility?*

K: COVID-19 was an accident that rolled back global development by a couple of years. This is a virus that is less than 1% lethal. Imagine what would happen if you raise the lethality rate by a factor of 10 or 20 or 50. It's debatable how quickly natural selection would favor something that is less lethal, but suffice it to say, it would not happen fast enough for humanity's liking.

Right now we don't know of any viruses that would cause new pandemics if released. But it's also true that at least 38,000 people can assemble an influenza virus from scratch. If people identify a new influenza virus that they think can cause a pandemic and share that information with the world, and if that pandemic could kill more than several million people (like COVID has), then you just gave 30,000 people access to an agent that is of nuclear-equivalent lethality.

A: *My understanding is that the U.S. Government is currently funding research programs to identify new potential pandemic-level viruses.*

K: Unfortunately, yes. The U.S. government thinks we need to learn about these viruses so we can build defenses — in this case vaccines and antivirals. Of course, vaccines are what have gotten us out of COVID, more or less. Certainly they've saved a ton of lives. And antivirals like Paxil are helping. So people naturally think, that's that's the answer, right?

But it's not. In the first place, learning whether a virus is pandemic capable does not help you develop a vaccine against it in any way, nor does it help create antivirals. Second, knowing about a pandemic-capable

virus in advance doesn't speed up research in vaccines or antivirals. You can't run a clinical trial in humans on a new virus of unknown lethality, especially one which has never infected a human — and might never. And given that we can design vaccines in one day, you don't save much time in knowing what the threat is in advance.

The problem is there are around three to four pandemics per century that cause a million or more deaths, just judging from the last ones — 1889, 1918, 1957, 1968 and 2019. There's probably at least 100 times as many pandemic-capable viruses in nature — it's just that most of them never get exposed to humans, and if they do, they don't infect another human soon enough to spread. They just get extinguished.

What that means is if you identify one pandemic-capable virus, even if you can perfectly prevent it from spilling over and there's zero risk of accidents, you've prevented 1/100 of a pandemic. But if there's a 1% chance per year that someone will assemble that virus and release it, then you've caused one full pandemic in expectation. In other words, you've just killed more than 100 times as many people as you saved.

A: *Is 1% your actual best guess of the chance that a newly identified zoonotic virus would be released with current technology?*

K: If identified as pandemic capable and from one of the families where virus assembly works, which is most of them, our current estimates range from between 0.5% and 3% per year. It's hard to judge because we know of only one historical example of a person who, if active today, definitely would do it, given access to the knowledge. That's Seiichi Endo of Aum Shinrikyo,[1] who was a graduate-level virologist out of Kyoto University. Aum wanted to obtain Ebola for use against civilians. Any graduate-trained virologist at Kyoto University today could assemble pretty much any of these viruses. Endo would have the skills — and the cults' budget certainly would have provided the resources. But honestly, it's so cheap these days that pretty much anyone with the relevant skill set makes enough money in their personal salary that they could afford the cost of the relevant reagents.

In addition to identifying pandemic-capable viruses, the other form of dangerous research is so-called gain of function, which is probably better termed virus transmissibility enhancement research. This is where scientists take viruses that are bad at transmitting human to human, but really good at killing you if they infect you, then try to engineer and evolve them to be more transmissible.

A: *My understanding is that we cannot point to a lot of concrete benefits to this kind of research. Is that correct?*

K: I cannot think of a single benefit from any kind of virus enhancement research or pandemic virus ID research. It's not important for developing

1. Aum Shrinrikyo (now called Aleph) is the Japanese doomsday cult responsible for the 1995 sarin gas attacks on the Tokyo subway. Endo led the group's attempts to produce botulinum and anthrax toxins.

vaccines and it has not been relevant to developing any antivirals. Nor has it focused attention or effort on the development of particularly effective countermeasures. Unless you have 100 million vaccine doses ready to go, something that spreads rapidly through the air traffic network is going to be too fast for us to get control of.

A: *That leads me to another question I had: a lot of your threat model seems to be about deliberate multisite release — someone releasing a virus in a bunch of airports, right?*

K: That's right. You could argue that that idea itself is an info hazard, but I struggle to believe that anyone capable of correctly assembling a virus would not think about releasing it in some place like an airport, presumably more than one airport.

I'm cynical enough to think that there are people like Seiichi Endo out there and that they're not just restricted to apocalyptic cultists. Certainly there are people like the Unabomber, who wanted to bring down the industrial system, which necessarily involves billions of people dying. This is someone who was good enough to become a mathematics professor at Berkeley. Would a modern day Ted Kaczynski study virology to learn how to manufacture a pandemic himself? Maybe.

A: *This seems to require a high level of logistical competence on the part of terrorists. If this is so feasible with current tech, why hasn't it happened yet? And why haven't we seen more than one credible attempt?*

K: The reason why we haven't seen any credible attempts with pandemic-capable viruses is we haven't had any pandemic-capable viruses to use.

We still don't know of any. The logistics of a "normal biological" attack — think of anthrax, botulism, tularemia — are difficult because you need to make a lot of it, purify it and disperse it over a large area. But that means it's more like a chemical weapon — it doesn't take advantage of biology's strength, which is self-replication.

So why hasn't it happened? The capability, thankfully, isn't there yet, but at some point it will be. And that's the hardest part: everything we do to try to keep this knowledge locked away is a matter of buying time. All we can do is delay. There are too many advances happening in too many different areas of biology to lock away that capability indefinitely. We're going to have to deal with a world where there are instructions for making pandemic agents that are accessible to researchers who can acquire the necessary DNA comprising the genome of that agent.

A: *Let's talk about delay, then. How do you think we can delay scientists from discovering these pandemic-capable agents?*

K: There are two ways that are the most promising.

Number one: we can find a way to make people liable for causing catastrophe. We can set the bar very high, say, something like 10 million deaths worldwide — direct or indirect — caused by some event for which you were clearly responsible. For example, if a scientist conducts research that is then used as a blueprint by somebody else, that would certainly qualify. Accidental releases would qualify too. Then you combine that with some sort of requirement for insurance, or even require general liability insurance to cover this. If institutions had to have insurance that factored in the potential negative externality cost of doing research on, say, viruses that could cause pandemics, their insurance premiums would be way higher than they are now. Then that means that governments, if they wanted to fund this kind of research, would have to throw a lot more money at it.

A: *So one approach is to make dangerous research more expensive. What's the second?*

K: The other form is more radical. The international community has agreed that nuclear weapons must never fall into the hands of non-state actors or terrorist groups. Pandemic viruses can kill more people than any nuclear weapon. Therefore, the same logic demands that we keep them out of the hands of terrorists.

If anyone credibly identifies a pandemic-capable agent, then they just handed it to tens of thousands of people. That's far worse than any degree of nuclear proliferation. Therefore, we can't allow that. We could enact a pandemic test ban treaty that specifically bans the laboratory experiments required to increase our confidence a given virus could cause a pandemic.

A: *You use this language of nuclear risk a lot — a test ban treaty, nuclear equivalent threats. I'm curious what you think about the extent that the lessons we've learned from nuclear threats are relevant to pandemic risk.*

K: So here's the really funny thing. The Asilomar Conference on Recombinant DNA[2] was held primarily for two reasons. Number one, the general public was afraid that recombinant DNA would lead to the next atom bomb. And scientists were not certain that recombinant DNA might create a fitness disadvantage that would allow it to spread in the wild and cause harm, particularly if applied to viruses that could cause pandemics. That caused molecular biologists to declare a moratorium on their own field until Asilomar, at which they concluded that, as best we can tell, we don't know how to create something that is fitness positive in the wild.

2. A group of roughly 140 scientists (as well as lawyers and physicians) met at The Asilomar Conference on Recombinant DNA in February 1975 to draft voluntary guidelines for research on recombinant DNA. The meeting set stringent standards to ensure research on recombinant DNA could proceed without endangering public health.

A: *It's an interesting contrast.*

K: That has held ever since then. But 30 years later, you have the editor in chief of *Science* essentially saying that the only way they weren't going to publish the genome of the 1918 influenza, which killed 50 million people, was if the federal government classified it.

A: *I'm thinking also of the 2014 gain-of-function moratorium and all of the pushback against that. It seems to me, from the outside, that there has been something of a culture shift in biology since the '70s. I'm wondering if you have thoughts on what caused that and what tools are available to help us shift back to a more security-conscious place.*

K: Honestly, I think it's going to be too late. I don't think the norms can change quickly enough. Even if they could, there are too many advances, and it's not often immediately clear how an advance can be misused. It's hard to turn down plausible new ways of saving people from cancer, heart disease or aging just because there is a chance that it might lead to another way to make pandemic class agents.

Historically, it's been difficult to say, "We need to change the rules now," because in the past science was always net positive. But now the risk of catastrophe is so high that we just can't afford to keep playing in the sandbox. Still, saying, "It's too dangerous and we need to stop" — that's a hard sell, especially for people who became scientists; the primary trait that drives them is curiosity. But there is a subset of threats where it does not matter how much you learn about it; you cannot counter it. And, unfortunately, pandemic-class agents appear to be in this category.

Learning more about the details of how these things work on a molecular level might well help us develop vaccines and antivirals. But vaccines and antivirals cannot help us contain a deliberately released pandemic. It doesn't matter if you can invent a perfect vaccine that is super easy to

make. You cannot manufacture and distribute faster than the pandemic is going to spread. There is just no way that biotech can help defend against catastrophic, deliberate pandemics, other than in diagnostics, figuring out where it is in order to try to tamp it down. Early warning is all you can do. What that means is that fighting pandemics and preparing for future pandemics and ensuring that that kind of event is something that we can reliably defend against is a job for physical scientists and engineers. It's a job for protective equipment. It's a job for germicidal light. It's a job for cryptographic methods of telling people how much risk they're at based on their connection network.

Right now most scientists are really not going to notice if the thing that they're working on happens to provide the key that will allow individuals to murder millions. USAID's DEEP VZN program had never considered the possibility that the viruses that they discover and post on their rank-ordered list by threat level would be misused.

A: *On that point — we've talked a lot about delaying pandemics, but we haven't spoken about Secure DNA yet.*

K: The basic idea of Secure DNA is this: the reagents required to assemble a virus are commonly available and cheap. There's no possible way of controlling them because they're required for basically all biomedical research, with one exception: in order to make a virus, you need the DNA encoding its genome. If we can prevent people from getting the DNA corresponding to particularly nasty viruses, then we can at least ensure that the risk from non-state actors is pretty minimal.

Now, this has been recognized for some time. A lot of folks rang the alarm bell on this way back in 2007. The leading companies then took it really seriously. It's really one of those shining examples of industry doing the right thing. The five leading gene synthesis providers at the time came together and decided they would screen orders for hazards and screen customers to make sure that they're legitimate people doing legitimate research. They formed what's called the International Gene Synthesis Consortium, and they claim that they screen about 80% of global synthetic DNA. They do it even though it costs them significant amounts of money to do that screening — as it requires expert biologists to take a look at all the false alarms that the screening algorithms throw up because a lot of biology is very similar to other biology.

The problem is that 80% is not 100%. So in terms of its effectiveness at actually preventing access to hazardous DNA, it definitely leaves something to be desired.

So we thought there had to be a better way to do screening than similarity search. How about we figure out the signature of a hazard in terms of exact sequences, calculate functional variance (that is, other sequences

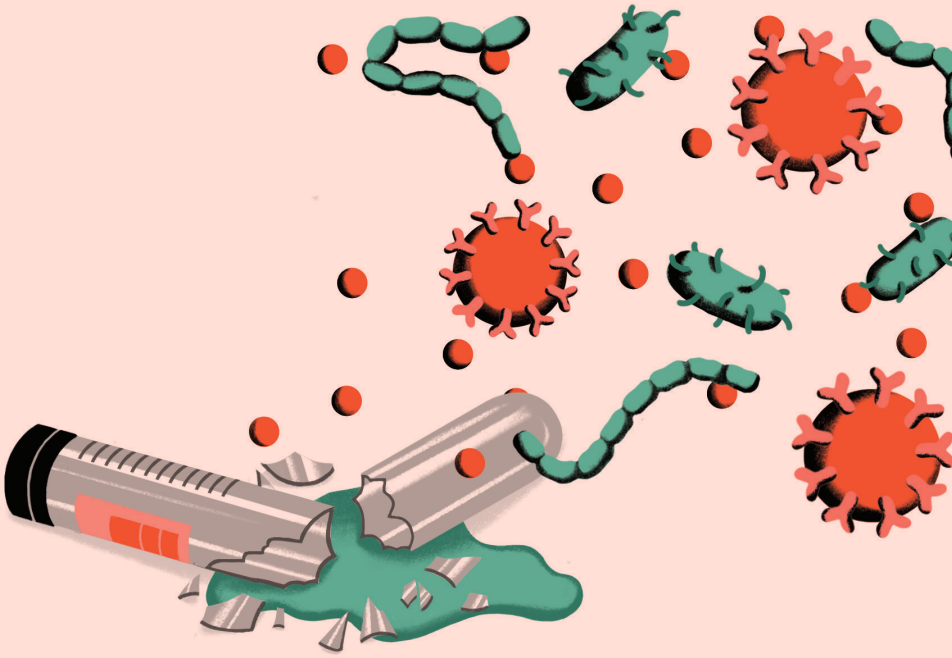
that could substitute for that particular signature), compare all of those to everything else ever sequenced, and throw out the sequences that match something unrelated to a hazard? Then all we need to do is check incoming orders for whether they match any of the fragments that define all of the things that we think are hazardous. That's way more computationally efficient and it doesn't raise these false alarm problems. That way we can screen orders without knowing what's in them, and we can also screen for hazards without knowing what they are.

That means that in principle, humanity could crowdsource threat identification. Instead of warning the world about a new threat, scientists who are very concerned about a particular way that biology could be used to cause harm could contact a curator of the secure DNA system and say, "I'm really worried about this." If the curator agrees, they can add it to the database along with a suitable number of decoys. Then synthesizers around the world would refuse to make that thing unless the ordering scientist had permission from their biosafety committee.

A: *And nobody ever has the full list of which viruses are dangerous.*

K: And no one learns other than the person who came up with the threat and the curator.

A: *Let's say we've delayed as long as possible and something gets through. Our best bet is to be able to detect it early, which I know is something else that you're working on.*



K: Suppose someone does find something nasty. Or suppose there's not something publicly known — some state biological weapons program that comes up with something nasty. What if it's like HIV? It's not obvious that it's spreading. You're not necessarily going to see things in the clinic any time soon. So how could we have detected something like HIV? Well, we know that it spread worldwide through the air traffic network. So what you want to do is monitor the air traffic network. The problem is you don't necessarily know what that hazard is going to look like.

So you have to look for some trait that is universal to threats. And when it comes to biology, all serious biological threats must be able to spread on their own, typically in an exponential growth pattern. So can you look for the pattern of exponential growth? The answer here is yes, you can, if you sequence all the nucleic acids that are present, and then look for specific sequences of unique fragments that have appeared, ideally across multiple monitoring sites. Then you can pull out those reads and say,

There is just no way that biotech can help defend against catastrophic, deliberate pandemics, other than in diagnostics — figuring out where it is in order to try to tamp it down.

"What is that? Does it look like it's a threat? Do we need to take action?" This will provide us with reliable early warning of any biological threat that's spreading human to human.

You can extend that approach to sequencing rivers. All the DNA in a watershed washes down into a river. That would allow you to detect things like gene drives in the environment as well. So the combination of air traffic network, untargeted metagenomic sequencing and environmental untargeted metagenomic sequencing would provide us with reliable early warning of anything threatening. It does not matter what it is. It does not matter if it's an extraordinarily competent adversary. It doesn't matter if it's a superhuman adversary. We will still be able to see it.

A: *All right, so we've detected it. And then what do we do then? How do we protect ourselves?*

K: You need to figure out how far it's spread, because exponential growth detection is not as sensitive as looking for a particular thing. But the CDC, for example, has finally gotten on the ball enough to have this wastewater-monitoring network for COVID strains across most American cities of

any decent size. Many other nations have similar sorts of networks. The next step, if we see something in the observatory, is to alert bodies like the DOD and CDC. The CDC can then tell all of its monitoring sites, "Look for this new thing. Here are some primers that you can use to amplify it and detect it." Then you figure out where it is in every town above, I don't know, 100,000 people. Then you drill down in the town where it's present, develop diagnostics and figure out who has it — assuming it's in people rather than something in the environment. Then you need to limit spread using standard anti-pandemic containment measures. Once we know the sequence of the hazard, which is what the observatory tells us, and we can design versions of these diagnostics that can sense it, we need the manufacturing capacity to scale those up really fast. In the meantime, we might need to do a lockdown in the cities that have it.

Your next question: what do we do to stop it? I'm speaking here from the perspective of an agent that could conceivably cause civilizational collapse. Suppose COVID had a 90% mortality rate. I can imagine people refusing to go out. That's good for curtailing the spread of the pathogen, but people need food and water and power at a minimum. We probably need law enforcement too — some kind of order.

Society could still function without health care in an extreme emergency. Many people would die without the health care system, yes, but we can do without it. But the people who are responsible for producing and distributing food, water and power absolutely must be willing to keep doing their jobs. That means that we have to give them good enough protective equipment.

So we need 30 million suits of protective equipment that requires zero training that can be delivered to them all within days, and that will reliably keep them from getting infected with anything that we think is nasty enough to warrant this kind of response.

A: *When you're imagining that next-generation totally reliable PPE, what does that look like?*

K: There's a couple of ways of doing it. The simplest version is a headpiece that ideally has complete plastic all the way across the front so you can see the face, covers the back of the head, and has some sort of clasp around the neck. It doesn't need to be very tight because you're creating positive pressure by pumping air through a HEPA filter into the inside. We can probably improve it by adding, say, LEDs that emit ultraviolet light to help sterilize the air going through. It needs to be comfortable. Ideally it needs to be stylish — you want as many people to be willing to wear it as possible, certainly in the early days. And it needs to be possible to take it off without self-contaminating and then infecting yourself. There also needs to be some way of sterilizing the equipment so that you can wear it again the next day — germicidal light is our best bet.

And that is our other best defense. Low-wavelength light between 200 and 230 nanometers is germicidal. It destroys viruses and bacteria, but it doesn't appear to hurt multicellular organisms because it's absorbed by proteins. Preliminary studies suggest even high exposures to this kind of light are safe. If we were to install these low-wavelength lights indoors, continuously and at a background level, under the current safety guidelines, it would reduce the amount of aerosolized pathogen in the air by 99% inside of five minutes. It could basically eliminate most aerosol- and contact-based transmission. What it wouldn't do is hit aerosols and the respiratory droplets from person-to-person transmission.

We also don't have generation mechanisms that could flick on and give us a higher dose. But if we can make LEDs that can do this, we could listen for two different voices in a room. When they're talking, the light switches to higher intensity. That could, in theory, inactivate the viruses before they move between two people in close conversation. We're not sure yet that this works, and we need to run comprehensive safety studies.

But it's incredibly promising because anything sufficient to prevent a serious future pandemic could probably also prevent the vast majority of the pathogens that infect us day-to-day. If we can actually harden our spaces to make them immune to transmission of pandemic viruses, then we've also just eliminated virtually all infectious disease. U.S. employers lose $300 billion a year to lost productivity from illness, specifically from infectious agents. That's $300 billion a year that could be saved.

Bluntly, we're not going to spend a lot of money on pandemic preparedness. No country in the world has. Maybe that will change if there's a deliberate attack first; people psychologically respond to attacks from other humans more than we do from natural catastrophes. I think the U.S. government has been unusually incompetent by failing to invest in pandemic preparedness, but that's the state of affairs for basically every nation in the world, with very few exceptions. But if we can address annual ongoing economic losses from standard infectious agents, that could convince people to install these things everywhere. Then we would be ready for the next pandemic.

A: *Since you mentioned the US government, I'm curious what you see as the biggest obstacles to implementing any of the ideas you've proposed. I'm also interested in how the response to the COVID pandemic has changed your views on what a response would realistically look like.*

K: The COVID pandemic has shown how difficult pandemic response will be if people don't believe there is a real threat. Imagine that the pathogen was something like HIV, which can circulate widely before anyone becomes symptomatic. Experts tell everyone about this new virus spreading across the globe that is like HIV and needs to be stopped. But no one has gotten sick yet. I think a lot of people would decline to believe the experts in that scenario.

The other lesson is that American institutions flat-out failed. I don't mean politically. I mean the CDC and the FDA themselves have arguably made the situation worse.

A: *Just to clarify, this is about speed to approve tests and vaccines?*

Kevin: Yes. The CDC and the FDA ensured that tests developed in many different universities simultaneously could not be used. There was the mess that was mask advising. The vaccines could have been approved faster if we had run challenge trials. I'm not confident in that assessment, but there is a significant probability that more Americans would be alive today if we had suspended the CDC and the FDA at the onset of the pandemic. On the whole, I think the FDA does a reasonable job of balancing benefits and risks for standard things, but in an emergency situation like a pandemic when you have to move fast — because every day you delay many thousands of people are dying — you just can't afford to have the same people governing the response. I don't think there's a human psyche on the planet that could manage that rapid flip. I would really like to see a separate system where power is formally transferred once an emergency is declared to people whose job it is to wait around and plan for emergencies.

A: *And on that cheerful note, is there anything else you'd like to say?*

K: I think the overall message has to be one of optimism. We still don't know of any capable agents. And it looks a lot like we can build technologies and launch them in plausible ways that don't necessarily require governments to respond or governments to take action using taxpayer dollars. It's possible that the bulk of the problem could be solved philanthropically, at least if you can get a few tens of billions of dollars. That's never been done before but it might be feasible in the wake of COVID, and if the tech can be proven to work.

We can build a world in which we don't have to fear the catastrophic misuse of biotechnology. And we have a road map for doing it, or at least we will soon. There's possibly more technologies that I missed or things that will be advanced. But we actually know that we have a problem and there is a clear and concrete set of definable potential solutions to that problem. There are multiple things that could solve the problem. Even if some of them don't work out, we'll still be OK. That's tremendously encouraging.

46

Rebuilding After the Replication Crisis

Stuart Ritchie

Over a decade has passed since scientists realized many of their studies were failing to replicate. How well have their attempts to fix the problem actually worked?

ILLUSTRATION BY
Adrian Forrow

An empty room with a large cardboard box in the center. A group of 102 undergrad students. They're split into three groups, and asked to sit either in the box, beside the box or in the room with the box removed. They complete a task that's supposed to measure creativity — coming up with words that link together three seemingly unrelated terms.

The results of this experiment? The students who sat *beside* the box had higher scores on the test than the ones *in* the box or those with no box present. That's because — according to the researchers — sitting next to the box activated in the students' minds the metaphor "thinking outside the box." And this, through some unknown psychological mechanism, boosted their creativity.

You might be laughing at this absurd-sounding experiment. You might even think I just made it up. But I didn't: It was published as part of a real study — one that the editors and reviewers at one of the top psychology journals, *Psychological Science*, deemed excellent enough to publish back in 2012.

To my knowledge, nobody has ever attempted to replicate this study — to repeat the same result in their own lab, with their own cardboard box. That's perhaps no surprise: After all, psychology research is infamous for having undergone a "replication crisis." That was the name that came to describe the realization — around the same time that the cardboard box study was published — that hardly any psychologists were bothering to do those all-important replication studies. Why check the validity of one another's findings when, instead, we could be pushing on to make new and exciting discoveries?

Developments in the years 2011 and 2012 made this issue hard to ignore. A Dutch psychology professor, Diederik Stapel, was found to have faked dozens of studies across many years, and nobody had noticed, in part because barely anyone had tried to replicate his work (and in part because it's really awkward to ask your boss if he's made up all his data). Psychologists published a provocative paper[1] that showed that they could find essentially any result they wished by using statistics in biased ways — ways that were almost certainly routinely used in the field. And one of those hen's-teeth replication attempts[2] found that a famous study from "social

1. Joseph Simmons et al, "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychological Science* 22, no. 11 (November 2011): 1359–66. https://doi.org/10.1177/0956797611417632.

2. Stéphane Doyen et al, "Behavioral Priming: It's All in the Mind, but Whose Mind?" *PLoS ONE* 7, no. 1 (2012): e29081. https://doi.org/10.1371/journal.pone.0029081.

priming," the same social psychology genre as the cardboard box study — in which merely seeing words relating to old people made participants walk more slowly out of the lab — might have been an illusion.

Similar stories followed. As psychologists got their act together and tried replicating one another's work, sometimes in large collaborations where they chose many studies from prominent journals to try to repeat, they found that approximately half the time, the older study wouldn't replicate (and even when it did, the effects were often a lot smaller than in the original claim). Confidence in the psychological literature started to waver. Many of those "exciting discoveries" psychologists thought they'd made were potentially just statistical flukes — products of digging through statistical noise and seeing illusory patterns, like the human face people claimed to see on the surface of Mars. Worse, some of the studies might even have been entirely made up.

The replication crisis, alas, applies to a lot more of science than just silly social psychology research. Research in all fields was affected by fraud, bias, negligence and hype, as I put it in the subtitle of my book *Science Fictions*. In that book, I argued that perverse incentives were the ultimate reason for all the bad science: Scientists are motivated by flashy new discoveries rather than "boring" replication studies — even though those replications might produce more solid knowledge. That's because for scientists, so much hinges on getting their papers published — particularly getting published in prestigious journals, which are on the lookout for groundbreaking, boundary-pushing results. Unfortunately, standards are so low that many of the novel results in those papers are based on flimsy studies, poor statistics, sloppy mistakes or outright fraud.

I think it's fair to predict with confidence that, were the cardboard box study to be repeated, the results would be different. It's the kind of study—based on tenuous reasoning about how language affects thought, with statistical tests that, when looked at in detail, are right on the very edge of being considered "statistically significant"—that would be a prime candidate for a failed replication, should anyone ever try. It's the kind of research that psychologists now look back on with embarrassment. Of course, a decade later we've learned our lesson, and definitely don't do unreplicable studies like that any more.

Right?

The problems of fraud, bias, negligence and hype in science aren't going away anytime soon. But we can still ask to what extent things have gotten *better*. Are researchers doing better studies — by any measure — than they were in 2012? Has anything about the perverse publishing dynamics changed? Have all the debates (what actually counts as a replication?), criticisms (are common statistical practices actually ruining science?), and reforms (should we change the way we publish research?) that have swirled around the idea of the replication crisis made science — in psychology, or indeed in any field — more reliable? Fundamentally, how much more can we trust a study published in 2022 compared to one from 2012?

If you jumped ten years forward in time from 2012, what would you notice that's different about the way science is published? Certainly you'd see a lot of unfamiliar terms. For instance, unless you were a clinical trialist, you likely wouldn't recognize the term "preregistration." This involves scientists planning out their study in detail before they collect the data, and posting the plan online for everyone to see (the idea is that this stops them "mucking about" with the data and finding spurious results). And unless you were a physicist or an economist, you might be surprised by the rise of "preprints" — working papers shared with

the community for comment, discussion and even citation before formal publication. These ideas come under the rubric of "open science," a term that in 2012 you might have heard of (it's been around since the 1980s), but that in 2022 is discussed almost everywhere.

You'd also notice a big rise in scientific papers discussing a "crisis," as well as all sorts of special issues and debate pieces dedicated to the idea of replicability. Like never before, many scientists are looking inward and questioning the reliability of their work. There are also telling patterns in the tools they're using. The Open Science Framework, a website where scientists can post their plans, share their data and generally make their whole research process more transparent, had somewhere near zero users in 2012, but by the end of 2021 had hit 400,000. The number of new files posted by those users, and the number of preregistrations, have also risen exponentially. In the past, a major barrier to being open and transparent with research was that it was really difficult to do so (how would you share your data, pre-internet?). It's still far from super easy, but the technology has substantially improved, and a great many scientists are signing up to use it.

You'd also notice that scientific publishers are changing. One of my formative experiences as a PhD student, in 2011, was submitting a replication study to the *Journal of Personality and Social Psychology*, only to be told that the journal did not publish replications under any circumstances (you might be thinking, "WTF?" — and we were too). Now at that very same journal and a host of others, replications are encouraged, as is a set of other "open" practices — sharing data, code and

materials, and pre-registering hypotheses and analyses before a study is carried out. Some journals now publish an article's peer reviews alongside its online version, so the whole process is on view — hopefully encouraging reviewers to put in more effort, and allowing us to see where things went wrong in cases where reviewers missed important flaws.

Over 300 journals across a variety of fields now offer the ultimate form of preregistered research, the "Registered Report," where it's not just that a plan is posted and then the study goes ahead, it's that peer reviewers review a study plan before the study happens. If the plan passes this quality control — and the reviewers might suggest all sorts of changes before they agree that it's a good study design — the journal commits to publish it, regardless of whether the results are positive or negative. This is a brilliant way of making sure that decisions about publication are made on the basis of how solid the design of a study is — not on the perceived excitement levels of its results.

These are all encouraging developments, and represent impressive progress in and of themselves. A scientist from 2012 would find a lot to be optimistic about in 2022 — at least on the surface. But the number of people *talking* about the crisis, debating open science or signing up to a website isn't what we really want to know. And journals *offering* ways to make science better isn't much use if nobody takes them up. Have these changes actually made the science better?

***

Given its life-or-death importance, it's no surprise that medicine has seen more intensive self-study, more meta-research, than any other field. Researchers at organizations like the Cochrane Collaboration have been beavering away, rating medical trials for their quality and how much they risk bias in their findings. To take just two examples, they check whether the participants in a trial could've found out if they were getting the real treatment or the placebo control — in which case the "blinding" of the study would've failed, expectation might play a role, and the results wouldn't be reliable. They also check whether the randomization of the study worked properly: If people are randomly assigned to groups, then there shouldn't be any big differences in health status, or background, between them before the study starts. If the randomization goes wrong, any results you find might be related to preexisting differences rather than to the treatment you're testing, and you'll draw the wrong conclusion.

If studies take extra care, they can reduce problems with blinding, randomization and a bunch of other bias-related problems that occur in medical trials,[3] and reduce the likelihood that they get

It's possible to go through the motions of "open science" without it really affecting your research or the way you behave—a problem that's increasingly been spotted as more researchers sign up to "open science" techniques.

spurious results. And according to a large-scale analysis of the overall trend in the quality of randomized medical trials from 1966 to 2018, the research has gotten better on average. Failures of randomization are rarer now; fewer studies have problems with blinding. Indeed, every metric they looked at has improved over the years, though there's still a long way to go — for example, 52% of trials in the period 2010–2018 still had problems with blinding. There also wasn't evidence of any kind of acceleration in quality over the past decade in particular.

So if you're reading a medical trial published recently — and many of us did this a lot during the pandemic — it is more likely to be better than one published in previous decades (though only a *little* better since 2012). A lot of that probably has to do with regulations on the way trials are planned and reported: Researchers in medical trials are forced to be transparent in a way that would be unrecognizable to scientists in other fields, whose research can effectively be entirely done in secret.

But nothing like this analysis has been done for any other field. Instead, we have to look at some *proxies* for quality. In psychology, one such proxy might be "adherence to open research": How much of the new replication-crisis-inspired reforms do they follow? Sadly, for this, all we have so far is a starting point:[4] Only 2% of psychology studies from 2014–2017 shared their data online, and just 3% had a preregistration. These numbers will undoubtedly rise in future surveys — but we don't have those surveys yet, so we don't know how much. As for Registered Reports, uptake by scientists has been slow, regardless of how many journals offer it as an option. Changing a whole culture — that, like any culture, has built up a great deal of inertia and skepticism about change — is hard — even if you have very good reason to do so.

Using adherence to open research as a proxy for research quality is complicated by the fact that it's possible to post a preregistration and then simply not follow it, or write it so vaguely that it doesn't constrain your "mucking about" in the intended way. Medical researchers might nod wearily here — it's been clear to them for years that scientists often dishonestly "switch" the outcome of their experiment, which they'd written down in their registration, to something else if their main outcome — pain levels, blood pressure measurements, depression ratings — didn't show the effect they wanted. It's also possible to post your data set online and have it be poorly annotated, or at worst completely incomprehensible. That's if the data is even present: A study from earlier this year found that, in studies where the authors wrote that they'd be happy to share their data on request, only 6.8% actually did so when emailed. In other words, it's possible to go through the motions of "open science" without it really affecting your research or the way you behave — a problem that's increasingly been spotted as more researchers sign up to these "open science" techniques. If you want to really make your research open, you have to actually *mean* it.

3. Aside from problems with blinding and randomization, these include factors like the quality of the measurement of the outcome (does the trial use good-quality instruments or well-validated questionnaires, or are they likely to produce noisy, hard-to-interpret data?) and the patterns of which data is missing (did people drop out of your trial in a "nonrandom" way — that is, did sicker people tend to quit the study faster? If so, your final results might be skewed). A fully detailed description of the huge number of ways that trials can be biased can be found on the website riskofbias.info.

4. Tom Hardwicke et al, "Estimating the Prevalence of Transparency and Reproducibility-Related Research Practices in Psychology (2014–2017)," *Perspectives on Psychological Science* 17, no. 1 (2012): 239–51. https://doi.org/10.1177/1745691620979806.

Another proxy for research quality is sample size. All else equal, bigger studies are usually better — so, are studies bigger nowadays? This is another way of asking about *statistical power*, the ability to detect effects if they're really there in your data. Studies with low-powered analyses — and usually this means studies that are too small — risk missing true effects and picking up on false ones.

In some fields, some types of analysis have undoubtedly become more powerful. Genetics is among the most obvious: After many years of failed "candidate gene" research, where small-sample research led the field badly astray, genetic studies now regularly reach sample sizes in the millions, and produce results that are replicable (even if their precise implications are still hotly debated). In brain-imaging research, too, there's an increasing awareness that to say anything sensible about how the brain relates to various behaviors, traits or disorders, we usually need sample sizes in the thousands. Happily, we now have what we need: Studies published in recent years have used resources like the UK Biobank or the ENIGMA Consortium, both with tens of thousands of brain scans, to come to more reliable conclusions. Alas, that has reemphasized that much of what was done in the past, in small-sample neuroimaging studies, was next to worthless.

Meta-research does show increasing sample sizes over time in neuroimaging as a whole; I'm certain that such a study would find the same in genetics. In other fields, it's less clear: There's some evidence, for instance, of a modest increase in sample size in personality psychology over time, and a recent preprint "cautiously" suggested that studies in political science have gotten bigger in recent years.

In other fields, though, all we have are starting points but no data on long-term trends. Almost uniformly, the starting point is one of very low power. That's true for psychology in general, clinical and sports and exercise psychology in particular, ecology, global change biology (the field that studies how ecosystems are impacted by climate change), economics, and political science. Other areas like geography have seen glimmers of a replication crisis but haven't yet collected the relevant meta-scientific data on factors like statistical power to assess how bad things are. We'll need a lot more meta-research in the future if we want to know whether things are getting better (or, whisper it, *worse*).

Even then, the mere knowledge that studies are, say, getting bigger shouldn't reassure us unless those studies are also becoming more *replicable* — that is to say, a closer approximation to reality. And although areas like psychology and economics have attempted to replicate dozens of experiments, there hasn't been time to make the same attempts to replicate newer studies or compare the replication rates over time. We likely won't see meta-research like this for a long time — and for some fields, a very long time. Witness how long it took the Reproducibility Project: Cancer Biology, a heroic attempt to replicate a selection of findings in preclinical cancer research, to finish its research: It began in 2013, but only just reported its final mixed bag of results in December 2021.

What about those papers that aren't just low quality, but are actively fraudulent or otherwise brought about by nefarious activities like plagiarism? A change from 2012 is that more papers are being retracted — removed from the scientific literature due to some major deficiency or error (it's a higher proportion of papers overall too). Not all retractions are due to deliberate rule breaking, of course — some are due to the discovery of honest mistakes, among other reasons (the Retraction Watch website covers each new retraction as it arises, and tries to ferret out the story behind it). But we can see it as a good thing that mistakes and falsehoods are

being actively dealt with more often than they were even a decade ago — even if, going by the number of papers that are flagged on error-checking websites like PubPeer, the frequency of retraction should be a lot higher.

***

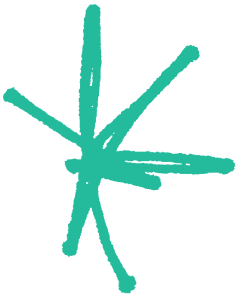
The economist Michael Clemens famously described the potential economic benefits of changes to immigration policy — removing restrictions and barriers to the movement of people across borders — as "trillion-dollar bills on the sidewalk." The benefits are just lying there, ready to be grabbed if politicians so choose. I think something similar is the case with science: Changes that would make dramatic improvements to the quality of research are *right there* — but, although they're often available, most scientists haven't even begun to pick them up.

And that's what's *really* different between now and a decade ago: We know a lot more about where science goes wrong, and we have a much longer list of potential tools to fix it. We've tried various reforms in several fields, producing useful lessons for other disciplines. We've developed technologies to improve the transparency of our research. And, in our more open, self-critical discussions (not to mention formal meta-research) about how science works, we've become much more aware of the hurdles — the incentives, the vested interests, the inertia and sometimes the sheer social awkwardness — that slow down the process of improving science.

But as you can see from my sorting through the scraps of evidence above, we have nowhere near the data we'd need to confidently argue that science is better now than a decade ago. Definitively answering this question will require substantially more meta-research across disciplines — and will likely require more reforms. The burst of meta-science that we *have* seen since the replication crisis mustn't be squandered: Pushing for the funding of much more such research should be a major priority for anyone who wants to improve science, and wants to do so using hard evidence.

Perhaps that evidence will tell us that our incremental fixes are ticking along nicely, steadily improving the quality of science as more and more researchers take up preregistration and open science and the rest. But equally, they might tell us that something more basic has to change. They might tell us that only a hard core of interested scientists are truly invested in "open science," and that the rest of the community needs to be incentivized into improving their own work. Perhaps — similar to the regulations for medical trials — we need simply to *require* that they follow a set of minimal standards before they receive funding. And maybe we need to fundamentally change how we approach science: a radical rethinking of the peer-review system, for example. Some have even argued that scientists' obsession with publication at all costs will only end if we get rid of scientific journals — or scientific papers themselves.

We shouldn't be afraid to trial and test new and creative ideas, even if they might make science look very different from the status quo a decade ago, or even today. That is, for science to become as trustworthy as we need it to be, it might — like those creative students back in 2012 — need to escape the cardboard box entirely.



54

Why Isn't the Whole World Rich?

Dietrich Vollrath

The question of why some countries join the developed world while others remain in poverty has vexed economists for decades. What makes it so hard to answer?

In 2019 there were about 648 million people living in extreme poverty, subsisting on the equivalent of $2.15 per day or less. Those 648 million people made up 8.4% of world population — representing an improvement over 1990, when 35.9% of people lived on that little. Yet even though extreme poverty has fallen, in 2018 about 80% of the world population still had material living standards less than one-third of that in the United States.[1]

One of the most frustrating things about the persistence of global poverty is that it is possible to eliminate it — at least within a country — in the space of a generation. In 1953, South Korea emerged from the Korean War desperately poor. It was almost entirely agrarian, and whatever infrastructure the Japanese had built during their occupation between 1910 and 1945 had been destroyed. In 1960 GDP per capita in South Korea was only around $1,200, lower than in Bangladesh, Nigeria or Bolivia, and about 6% of the GDP per capita in the United States.[2]

Shortly thereafter, everything started to change. In 1968 the growth rate of GDP per capita in South Korea topped 10%. Throughout the 1970s, per capita GDP grew nearly 9% each year on average, slowing only slightly through the 1980s and 1990s. By 1995, South Korean GDP per capita had eclipsed Portugal's. By 2008, it was ahead of New Zealand's and just behind Spain's. In 2020, GDP per capita in South Korea was nearly equal to that in the U.K. Not only is South Korea no longer developing; in many areas, it leads among developed nations.

What happened in South Korea offers proof that fundamental transformations of living standards are possible in a few decades. South Korea's experience, and similar

1. Based on GDP per capita. My calculations from Robert C. Feenstra, Robert Inklaar, and Marcel P. Timmer, *Penn World Table*, V10.0 (updated June 18, 2021), distributed by Groningen Growth Development Centre, https://doi.org/10.15141/S5Q94M.

2. GDP per capita data from the Penn World Table is adjusted for inflation and differences in the cost of living between countries. Other methods of estimation report somewhat different figures.

growth trajectories in Taiwan and Singapore, have often been referred to as an "economic miracles." But what if South Korea's economic growth wasn't something mysterious or unpredictable, but rather something that we could comprehend and, most importantly, replicate? At current rates of growth, living standards in the poorest countries in the world will eventually catch up to the United States — in about 700 years.[3] If we could identify what caused South Korea's takeoff, we might be able to make the miraculous seem routine, and see more countries catch up over decades and not centuries.

Economists have been engaged in research for decades to understand what happened in South Korea and other countries that left extreme poverty behind. It turns out to be one of the trickiest questions in economics. On the surface, it seems like the answer should be obvious: "Do whatever South Korea did." Or, more broadly, "Do whatever countries that grew rapidly did." But what, exactly, *did* South Korea do? And if we know, is it plausible to replicate it?

Scratching the Surface

Some of the first attempts to explain what happened in places like South Korea examined the role of "factors of production," as economists like to call them. Those factors include physical capital — tangible products like buildings, infrastructure and manufacturing equipment — and human capital — skills and education embodied in workers. In a famous and widely cited study, Greg Mankiw, David Romer and David Weil looked at how the accumulation of both factors was associated with economic growth.[4] Countries that allocated a large share of GDP toward producing new physical capital or had high levels of secondary school enrollment tended to grow faster than others. In addition, countries with lower population growth rates tended to grow faster, as they were able to equip each worker with more physical capital, raising their productivity.

Mankiw, Romer and Weil studied a broad set of nearly 100 countries from a very high level. Alwyn Young took a similar approach but narrowed his focus to four East Asian economies — Taiwan, South Korea, Hong Kong and Singapore — that had all experienced rapid economic growth.[5] What he found corroborated Mankiw, Romer and Weil's findings on physical capital to some extent. Young, however, attributed even more power to the changes in human capital. In each of the four countries, he found that families were having fewer children and investing more in their education. Increases in educational attainment created a more skilled workforce — an impact which Young was able to track

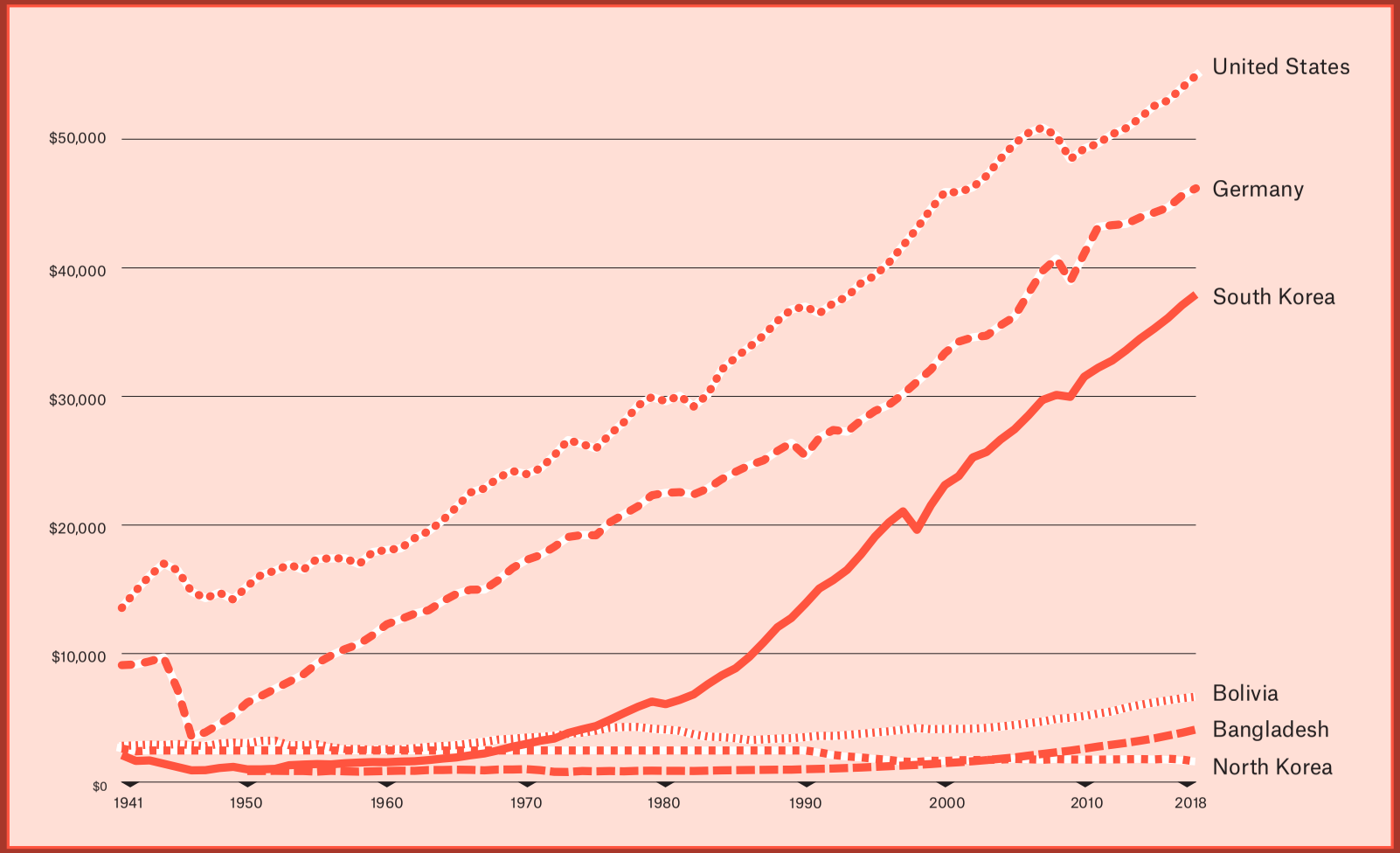
3. Calculation based on results from Dev Patel, Justin Sandefur, and Arvind Subramanian, "The New Era of Unconditional Convergence," *Journal of Development Economics* 152 (September 2021): 1–18, 102687, https://doi.org/10.1016/j.jdeveco.2021.102687.

4. N. Gregory Mankiw, David Romer, and David N. Weil, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* 107, no. 2 (May 1992): 407–437, https://doi.org/10.2307/2118477.

5. Alwyn Young, "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience," *Quarterly Journal of Economics* 110, no. 3 (August 1995), https://doi.org/10.2307/2946695.

GDP per capita, 1941 to 2018

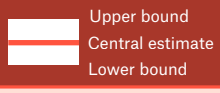
GDP per capita adjusted for price changes over time (inflation) and price differences between countries measured in international-$ in 2011 prices.

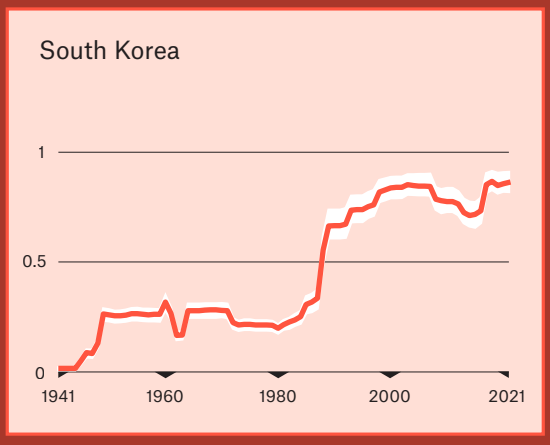


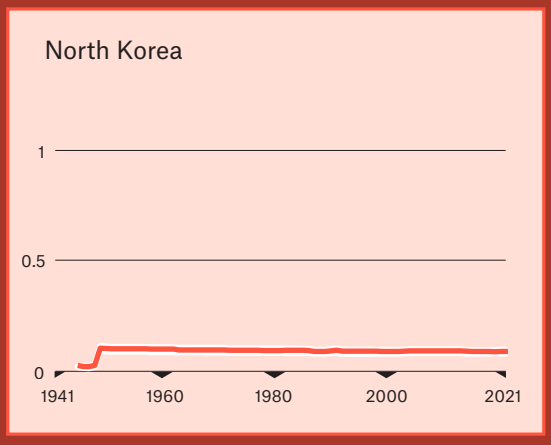
Source: Maddison Project Database 2020 (Bolt and van Zanden, 2020) / OurWorldInData.org/economic-growth

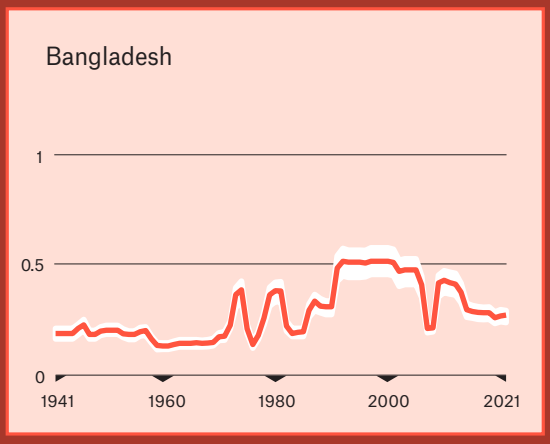
Electoral democracy, 1941 to 2021

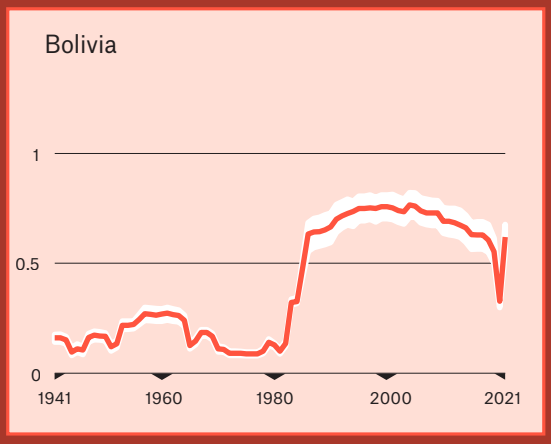
Based on the expert assessments and index by V-Dem. This index captures to which extent political leaders are elected under comprehensive voting rights in free and fair elections, and freedoms of association and expression are guaranteed. It ranges from 0 to 1 (most democratic).

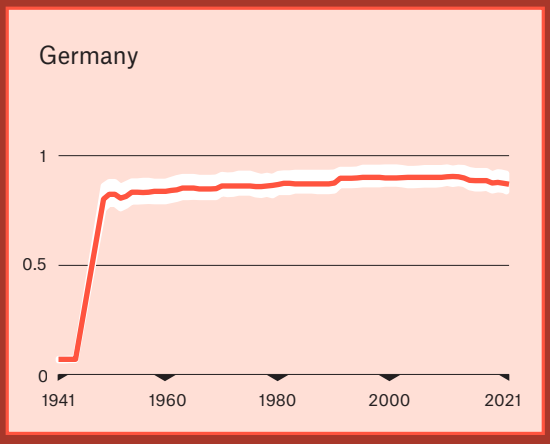


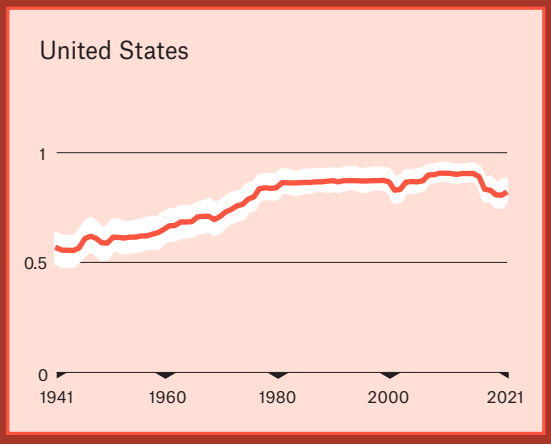












Source: OWID based on V-Dem (v12) / OurWorldInData.org/democracy

in more detail than Mankiw, Romer and Weil. Their slower population growth was associated with increased labor force participation by women and an increase in the share of the population that was of working age.

Research like this established *how* economic growth was able to accelerate in some countries, but it does not tell us *why* those changes took place in the first place. Why did capital formation speed up in South Korea or Taiwan (and not in Bangladesh or Nigeria)? Why did families start to have fewer, better-educated children in those same places?

What we are after is a deeper set of fundamental characteristics, policies and events that created the circumstances under which rapid economic growth occurred.

Institutions as Fundamentals

The hunt for the fundamental whys of rapid economic growth arguably defines the study of economics. Adam Smith was concerned with exactly this question in *The Wealth of Nations*. While that hunt has always been near the core of the discipline, there was an eruption of research on the subject in the decades following the studies by Young and Mankiw, Romer and Weil.

Within that literature, economists have tended to group those fundamentals of economic growth into three broad categories: culture (e.g., the willingness to trust and engage in trade with strangers), geography (e.g., ease of transportation) and institutions (e.g., security of property rights). Of the three categories, institutions have received the most attention. This is in part because they tend to be more legible to economists than issues of geography or culture, and in part because they would appear to be more amenable to change.[6]

But what exactly is an institution? Douglass North, the Nobel Prize winner credited with originating the study of institutions as a driver of long-run growth, has defined them as "humanly devised constraints that structure political, economic, and social interactions."[7] That is so broad it offers little chance of identifying real policies or changes that countries could pursue. Researchers who took North's ideas and ran with them contributed in part by being more specific. In early work, Daron Acemoglu, Simon Johnson and James Robinson, responsible for initiating detailed empirical research into institutions, focused on the security of private property rights, measured by either the risk of expropriation (based on

6. Implicitly, there is a fourth fundamental to add to this list: luck. It might be that some of what explains growth in South Korea or other economic successes is a fortunate set of contingent circumstances, and there isn't any way to make a miracle. Even if I had a complete physical and psychological profile of Serena Williams, I probably cannot make complete sense of her dominance, which at times might have benefited from a favorable draw or a lucky bounce.

7. Douglass C. North, "Institutions," *Journal of Economic Perspectives* 5, no. 1 (Winter 1991): 97, https://doi.org/10.1257/jep.5.1.97. The work summarized in that article originates in Douglass C. North and Robert Paul Thomas, The Rise of the Western World: A New Economic History (Cambridge: Cambridge University Press, 1973), and Douglass C. North, *Structure and Change in Economic History* (New York: Norton, 1981).

assessments by investors) or the legal constraints on government executives (based on assessments by political scientists).[8]

Work by Acemoglu, Johnson and Robinson, and those that followed, looked across a wide set of countries, searching for common institutional elements that existed in all the countries that experienced rapid economic growth (or that were absent in those that did not). These studies focused at first on measurements of institutions and growth during the 20th century, but soon incorporated data from even earlier. The same three authors (along with Davide Cantoni) studied the importance of an institution we could call "equality before the law" by examining the effect of Napoleonic reforms made in Germany at the turn of the 19th century on subsequent development.[9] In other work, they estimated that European countries with more representative institutions, like Britain and the Netherlands, were able to grow more quickly in response to the opening of trans-Atlantic trade routes than absolute monarchies like Spain and Portugal.[10]

These authors and the literature that followed them tended to find that things like robust property rights for individuals and governments with clear restraints on executive power, democratic political processes and a lack of government corruption were all associated with economic growth.

Those institutions certainly sound "right." They are things we'd associate with almost any major developed country like the U.S., France or Germany. But, at heart, most of these studies share the same fundamental issue as those that looked at capital accumulation: Just because certain institutions are present in places that had rapid economic growth, that doesn't mean they were necessary for the miracle to occur. Perhaps things like property rights and a lack of corruption are "luxury goods" that rich countries can afford to indulge in but are not, in fact, the reason those countries became rich?

The problem gets even thornier when researchers try to pin down how to even measure an "institution" in the first place.

A concrete example: The World Bank has a set of "Governance Indicators" it collects from each country. Those indicators include a measure of the "control of corruption" that a country has. For example, in 2020 Eritrea had a "control of corruption" indicator of −1.33, quite low. Mauritius had a 0.47, which is around the middle of the pack, and Denmark had a 2.27, among the highest. In

8. Daron Acemoglu, Simon Johnson, and James A. Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review* 91, no. 5 (December 2001): 1369–1401, https://doi.org/10.1257/aer.91.5.1369. An accessible introduction to their body of work is Daron Acemoglu and James A. Robinson, *Why Nations Fail: The Origins of Power, Prosperity, and Poverty* (New York: Crown Publishers, 2012).

9. Daron Acemoglu, Davide Cantoni, Simon Johnson, and James A. Robinson, "The Consequences of Radical Reform: The French Revolution," *American Economic Review* 101, no. 7 (December 2011): 3286–3307, https://doi.org/10.1257/aer.101.7.3286.

10. Daron Acemoglu, Simon Johnson, and James A. Robinson, "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth," *American Economic Review* 95, no. 3 (June 2015): 546–579, https://doi.org/10.1257/0002828054201305.

terms of absolute ranking, it is probably correct that Eritrea is more corrupt than Mauritius and that both are more corrupt than Denmark.

But do the numbers themselves mean something? Is Denmark exactly 4.8 times less corrupt than Mauritius? If Eritrea managed to raise their index to –1, would that imply the same change in corruption as Mauritius moving to 0.80? The answer to both questions is obviously no. At best the numbers let us rank countries on these dimensions of governance, but there is no sense that 2.27 means anything in practice.

The statistical analysis that establishes the link between control of corruption and economic growth assumes, however, that the corruption index has a precise numerical meaning.[11] It's not that the statistical analysis is wrong — it's that it has no practical interpretation. The control-of-corruption index, like other World Bank governance indicators, is based on survey data. But people in rich countries are more likely to give their institutions high ratings. In one striking case, Edward Glaeser et al. pointed out that Singapore has historically scored highly on measures like constraint on executive power — even when it was ruled by Lee Kuan Yew, a dictator who had no constraints on his power but did happen to respect property rights.[12] Ideally, economists would try to control for confounding variables like wealth or education, but the fact that there are only about 50 to 70 countries with available data makes that impossible. As a result, the measures are circular: They tell us that Denmark is better governed than Mauritius or Eritrea, but not much else.

Just because certain institutions are present in places that had rapid economic growth, that doesn't mean they were necessary for the miracle to occur. Perhaps things like property rights and a lack of corruption are "luxury goods" that rich countries can afford to indulge in but are not, in fact, the reason those countries became rich?

11. Paolo Mauro, "Corruption and Growth," *Quarterly Journal of Economics*, 110, no. 3 (August 1995): 681–712, https://doi.org/10.2307/2946696.

12. Edward L. Glaeser, Rafael La Porta, Florencio Lopez-De-Silanes, and Andrei Shleifer, "Do Institutions Cause Growth?," *Journal of Economic Growth* 9, no. 3 (September 2004): 271–303, https://doi.org/10.1023/B:JOEG.0000038933.16398.ed.

This isn't a problem unique to measuring the degree of corruption. Every index of institutional quality is subject to this critique, because every index is attempting to assign numbers to something that is not inherently quantifiable: the degree of democracy, the rule of law, government effectiveness, respect for property rights, etc. In each case, the research might indicate that "being like Denmark" is a good thing, without any practical way of expressing what that means.

Experimenting With History

The picture I painted of cross-country research on economic growth is bleak, but those issues are not lost on researchers. Knowing these issues, scholars have tried to establish better evidence for which institutions matter for economic growth.

Much of this research is based on an examination of historical or natural experiments. Once again, South Korea is a useful example. After World War II, the Korean peninsula was, of course, partitioned between South and North Korea. The two countries share similar geography, so the miracle in South Korea and the utter lack of one in North Korea cannot be attributed to their endowment of minerals or physical access to foreign markets. They have a shared language and culture, so it is hard to say that there was something unique about the South Korean culture or history that prompted the miracle there (or halted it in North Korea). They both were left devastated and poor by the Korean War.

What's left as an explanation is that the set of institutions governing economic activity in the two countries were distinct after 1953. The North adopted a communist ideology and built a set of economic institutions around it. We can see the results of that today. North Korea has failed, by any plausible metric, to advance economically. In addition to the lack of individual freedom, living standards are among the worst in the world, and North Korea continues to suffer from recurring issues such as famine that advanced economies like South Korea left behind years ago.

This example is useful in that it tells us institutions matter for economic growth, and unlike other research can more clearly eliminate other options like geography or culture. It also doesn't require us to assign an artificial index to the institutions of South Korea or North Korea. We know they're different, and that's enough.

What that case study lacks, of course, is a clear answer to *which* institutions were the relevant ones making South Korea an economic miracle. Was it the subsidization of the "chaebol" — conglomerates like Samsung, Hyundai or LG — with cheap credit? Was it, uncomfortably, the lack of real democracy until 1988? Was it the promotion of exports versus domestic consumption? We can't know from this simple comparison.

Research has thus continued to search for more historical natural experiments where the nature of a particular institution is much more apparent. The experiments the authors rely on are often quite clever. Melissa Dell compared areas of Peru subject to a Spanish forced-labor requirement called the "mita" to areas that were not and found that they have lower living standards centuries later.[13] Lakshmi Iyer found that areas of India subject to direct British rule (as opposed to those ruled through proxies) have lower investments in

schooling and health today.[14] Stelios Michalapoulos and Elias Papaioannou compared areas of sub-Saharan Africa that had historically more sophisticated political structures prior to colonization continue to be richer today than areas that were less organized.[15] In each case, a very specific institution — a forced labor regime, British Raj ended decades ago, the Spanish forced-labor system in Peru ended over two centuries ago, and the historical political organization of sub-Saharan Africa are just that — historical. What we learn from these studies is that institutions can have persistent effects well after the institution disappears, implying that countries or regions can get stuck in a poverty trap. Once the region is impoverished, it's more likely to stay poor.

What we learn from these studies is that institutions can have persistent effects well after the institution disappears, implying that countries or regions can get stuck in a poverty trap. Once the region is impoverished, it's more likely to stay poor.

direct British rule, precolonial political structure — was found to have a significant effect on contemporary economic outcomes.

The empirical work here is on more solid ground, and the authors avoid the measurement issues mentioned above. But these studies, by narrowing their focus to specific historical experiments and individual institutions, have their own limitations. These studies don't tell us about the immediate effect of any of these institutions. The countries or regions can get stuck in a poverty trap. Once the region is impoverished, it's more likely to stay poor.

These papers work as cautionary tales; they tell us what won't work, but not what will work. And while they don't provide any silver bullets for generating economic growth, they remain valuable contributions to the study of development. This work is eliminating bad options from the menu of institutional choices that countries could make.

13. Melissa Dell, "The Persistent Effects of Peru's Mining *Mita*," *Econometrica* 78, no. 6 (November 2010): 1863–1903, https://doi.org/10.3982/ECTA8121.

14. Lakshmi Iyer, "Direct versus Indirect Colonial Rule in India: Long-Term Consequences," *The Review of Economics and Statistics* 92, no. 4 (November 2010): 691–713, https://doi.org/10.1162/REST_a_00023.

15. Stelios Michalapoulos and Elias Papaioannou, "Pre-colonial Ethnic Institutions and Contemporary African Development," *Econometrica* 81, no. 1 (January 2013): 113–152, https://doi.org/10.3982/ECTA9613.

Negotiating for Growth

Alongside the literature on what not to do, there is recent work that attempts to be more constructive. Acemoglu and Robinson, who helped initiate the empirical study of institutions, are among the leaders in this new line of inquiry as well.[16] The key here is a change in the question. Rather than asking what the right institutions are to promote growth, they ask why failed institutions persist. For them, countries stagnate at low levels of development because there is a stalemate among interest groups; despite the aggregate benefit, no group is willing to implement an improved set of institutions.

What their research suggests is that breaking out of that stalemate requires a fundamental expansion of the distribution of economic and political power within a country. By incorporating more people in economic and political decision-making, they argue, a country is better able to negotiate a set of economic institutions that promote economic development.

This sounds promising, but can we see it in the data? These authors and others have made progress and are beginning to provide supportive empirical work. What sets them apart from earlier work is that they have the benefit of knowing that mistakes were made in the past. A good example is from Acemoglu and Robinson along with coauthors Suresh Naidu and Pascual Restrepo.[17] They show that the transition to democracy leads to higher economic growth in the future, finding GDP per capita is around 20%

This result is exciting, in part because it suggests that something inherently positive — wider representation and democracy — is also conducive to economic growth. But it doesn't mean we've cracked the code and are capable of generating economic miracles at will.

16. Daron Acemoglu and James A. Robinson, "Political Losers as a Barrier to Development," *American Economic Review*. 90(2): 126–130, https://doi.org/10.1257/aer.90.2.126. This is also illustrated further in Acemoglu and Robinson, *Why Nations Fail*.

17. Daron Acemoglu, Suresh Naidu, Pascual Restrepo, and James A. Robinson, "Democracy Does Cause Growth," *Journal of Political Economy* 127, no. 1 (February 2019): 47–100, https://doi.org/10.1086/700936.

18. Consistent with the Acemoglu, Naidu, Restrepo, and Robinson (2019) findings, South Korea did eventually democratize in 1988 and now enjoys living standards roughly equal to those in Western Europe. China, on the other hand, has failed to expand political representation and its own growth miracle is already showing signs of slowing down well short of reaching that level of GDP per capita.

higher in a democracy compared to an otherwise identical nondemocracy. What they see is that countries that democratize invest significantly more in public health and education, consistent with the initial work that Mankiw, Romer and Weil and Alwyn Young did on economic growth.

They explicitly take on all of the empirical issues I complained about above. They do not try to quantify "democracy" along some arbitrary scale (e.g., North Korea is a one, the U.S. is a seven, etc.). They instead focus on a simple comparison of places that clearly democratized versus those that did not. They use several methods to try to assure themselves, and us, that their results are coming from the causal effect of democracy on growth, and not the other way around. This includes a sort of natural experiment where democratization is more likely to occur when more neighboring countries are democracies.

Some counterexamples may immediately come to mind. South Korea, whose economy took off in the '60s, did not democratize until 1988, and China has undergone impressive economic growth without democratizing at all. But once Acemoglu, Naidu, Restrepo and Robinson make the comparison across all countries, it turns out that their experiences are something of an outlier, not the norm. And in both, there were events that led to a widespread expansion of the distribution of economic power, even though it was not accompanied by political power: the massive redistribution of land in South Korea following World War II and the market reforms in the 1970s and '80s in China that gave more people rights over their land and assets.[18]

This result is exciting, in part because it suggests that something inherently positive — wider representation and democracy — is also conducive to economic growth. But it doesn't mean we've cracked the code and are capable of generating economic miracles at will. Countries that do expand the distribution of political and economic power still have to negotiate the institutions supporting growth. This is where our expanding knowledge of which institutions don't work becomes valuable, helping eliminate dead ends.

Making Modest Conclusions

At this point the situation may seem rather grim. Can we say, with any confidence, that we know the set of policies or institutions that can create the rapid economic growth seen in South Korea and others? The frank answer is no.

But this does not mean we are at a complete loss. Do not dismiss the power of the cautionary tales I mentioned. While the Korean "experiment" didn't tell us what exactly South Korea did right, it continues to provide a vivid lesson that the North Korean centrally-planned authoritarian regime was not a viable economic path to take. Documenting which institutions don't work is slow, but it is progress nonetheless. Furthermore, recent results regarding the importance of the distribution of economic and political power mean we understand more about the conditions that can cause good institutions to arise.

Can we make an economic miracle? No. Do we understand what might make economic miracles more likely? To some extent, yes. That wishy-washy answer doesn't sound very inspiring, but it represents a tremendous amount of progress. The series of critiques and incremental improvements I've described is an example of the research process at work. Given the stakes, the slow pace is frustrating, but we are headed in the right direction.

66

Is Wine Fake?

Scott Alexander

Wine commands wealth, prestige, and attention from aficionados. How much of what they admire is in their heads?

Your classiest friend invites you to dinner. They take out a bottle of Chardonnay that costs more than your last vacation and pour each of you a drink. They sip from their glass. "Ah," they say. "1973. An excellent vintage. Notes of avocado, gingko and strontium." You're not sure what to do. You mumble something about how you can really taste the strontium. But internally, you wonder: Is wine fake?

A vocal group of skeptics thinks it might be. The most eloquent summary of their position is *The Guardian's* "Wine-Tasting: It's Junk Science," which highlights several concerning experiments:

> In 2001 Frédérick Brochet of the University of Bordeaux asked 54 wine experts to test two glasses of wine – one red, one white. Using the typical language of tasters, the panel described the red as "jammy' and commented on its crushed red fruit.
>
> The critics failed to spot that both wines were from the same bottle. The only difference was that one had been coloured red with a flavourless dye.

And:

> In 2011 Professor Richard Wiseman, a psychologist (and former professional magician) at Hertfordshire University invited 578 people to comment on a range of red and white wines, varying from £3.49 for a claret to £30 for champagne, and tasted blind. People could tell the difference between wines under £5 and those above £10 only 53% of the time for whites and only 47% of the time for reds. Overall they would have been just as successful flipping a coin to guess.

Wikipedia broadly agrees, saying:

> Some blinded trials among wine consumers have indicated that people can find nothing in a wine's aroma or taste to distinguish between ordinary and pricey brands. Academic research on blinded wine tastings have also cast doubt on the ability of professional tasters to judge wines consistently.

But I recently watched the documentary *Somm*, about expert wine-tasters trying to pass the Master Sommelier examination. As part of their test, they have to blind-taste six wines and, for each, identify the

grape variety, the year it was produced, and tasting notes (e.g., "aged orange peel" or "hints of berry"). Then they need to identify where the wine was grown: certainly in broad categories like country or region, but ideally down to the particular vineyard. Most candidates — 92% — fail the examination. But some pass. And the criteria are so strict that random guessing alone can't explain the few successes.

So what's going on? How come some experts can't distinguish red and white wines, and others can tell that it's a 1951 Riesling from the Seine River Valley? If you can detect aged orange peel, why can't you tell a $3 bottle from a $30 one?

In Vino Veritas

All of those things in *Somm* — grape varieties, country of origin and so on — probably aren't fake.

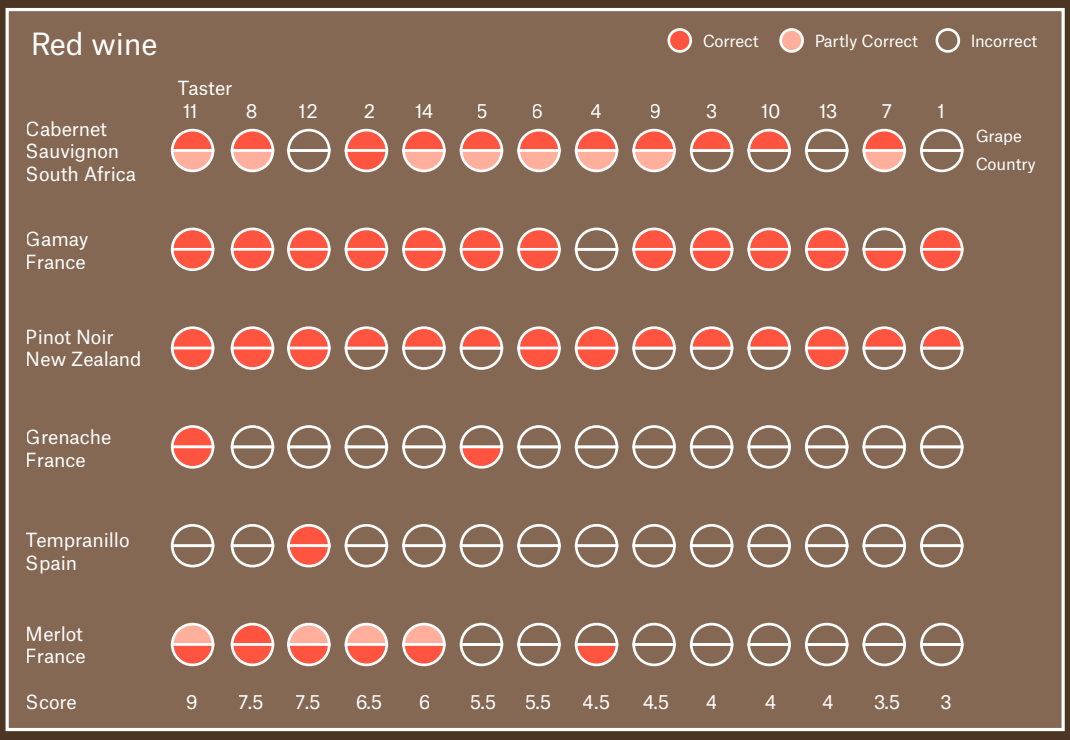
The most convincing evidence for this is "Supertasters Among the Dreaming Spires," from 1843 magazine (also summarized in The Economist). Here a journalist follows the Oxford and Cambridge competitive wine-tasting teams as they prepare for their annual competition. The Master Sommelier examination has never made its results public to journalists or scientists — but the Oxbridge contest did, confirming that some of these wine tasters are pretty good. Opposite is The Economist's graphical summary.

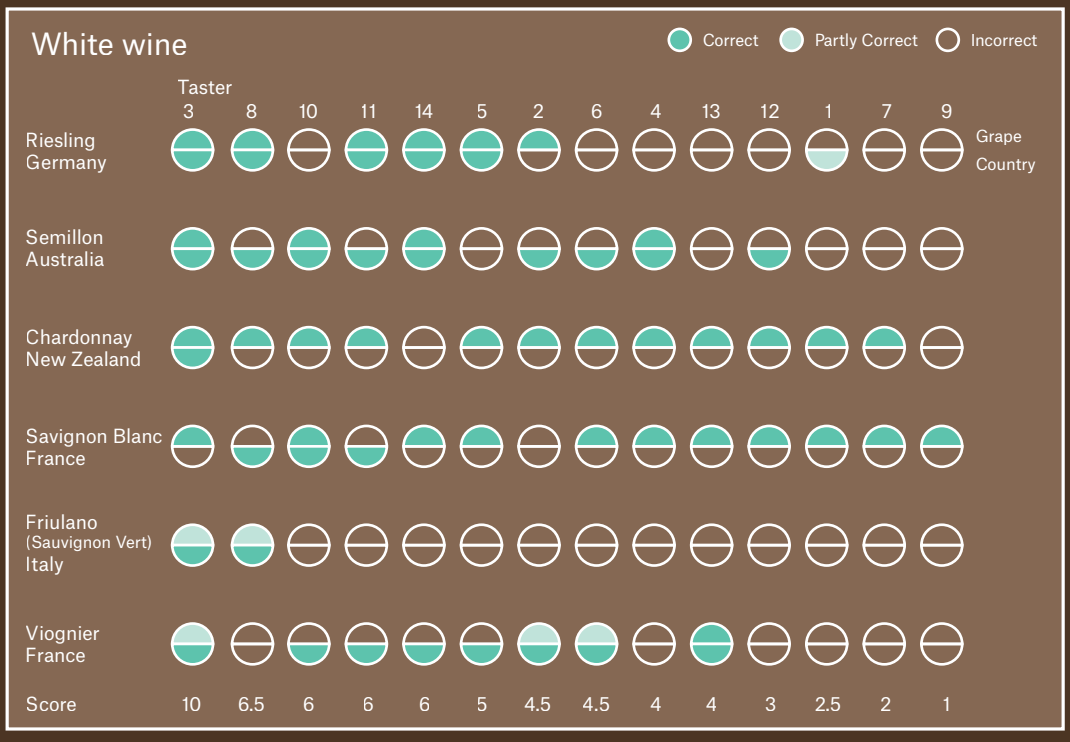
Top scorers were able to identify grape varieties and countries for four of the six wines. In general, tasters who did well on the reds also did well on the whites, suggesting a consistent talent. And most tasters failed on the same wines (e.g., the Grenache and Friulano), suggesting those were genuinely harder than others.

If the Oxbridge results are true, how come Brochet's experts couldn't distinguish red and white wine? A closer look at the original study suggests three possible problems.

First, the experts weren't exactly experts. They were, in the grand tradition of studies everywhere, undergraduates at the researchers' university. Their only claim to expertise was their course of study in enology, apparently something you can specialize in if you go to the University of Bordeaux. Still, the study doesn't say how many years they'd been studying, or whether their studies necessarily involved wine appreciation as opposed to just how to grow grapes or run a restaurant.

Results of the Oxford-Cambridge Varsity blind-tasting match
February 15, 2017





Source: Economist.com

Second, the subjects were never asked whether the wine was red or white. They were given a list of descriptors, some of which were typical of red wine, others of white wine, and asked to assign them to one of the wines. (They also had the option to pick descriptors of their own choosing, but it's not clear if any did.) Maybe their thought process was something like "neither of these tastes red, exactly, but I've got to assign the red wine descriptors to one of them, and the one on the right is obviously a red wine because it's red colored, so I'll assign it to that one."

Third, even if you find neither of these exculpatory, tricking people just works really well in general. Based on the theory of predictive coding, our brains first figure out what sensory stimuli *should* be, then see if there's any way they can shoehorn actual stimuli to the the expected pattern. If they can't, then the brain will just register the the real sensation, but as long as it's *pretty close* they'll just return the the prediction. For example, did you notice that the word "the" was duplicated three times in this paragraph? Your brain was expecting to read a single word "the," just as it always has before, and when you're reading quickly, the mild deviation from expected stimuli wasn't enough to raise any alarms.

Or consider the famous Pepsi Challenge: Pepsi asked consumers to blind-taste-test Pepsi vs. Coke; most preferred Pepsi. But Coke maintains its high market share partly because when people are asked to *nonblindly* taste Coke and Pepsi (as they always do in the real world) people prefer Coke. Think of it as the brain combining two sources of input to make a final taste perception: the actual taste of the two sodas and a preconceived notion (probably based on great marketing) that Coke should taste better. In the same way, wine tasters given some decoy evidence (the color of the wine) combine that evidence with the real taste sensations in order to produce a conscious

perception of what the wine tastes like. That doesn't necessarily mean the same tasters would get it wrong if they weren't being tricked.

Pineau et al.[1] conducted a taste test that removed some of these issues; they asked students to rank the berry tastes (a typical red wine flavor) of various wines while blinded to (but not deceived about) whether they were red or white. They were able to do much better than chance ($p<0.001$).

The Price Is Wrong

Just because wine experts can judge the characteristics of wine doesn't mean we should care about their assessments of quality. Most of the research I found showed no blind preference for more expensive wines over cheaper ones.

Here my favorite study is Goldstein et al.,[2] "Do More Expensive Wines Taste Better? Evidence From a Large Sample of Blind Tastings." They look at 6,175 tastings from 17 wine tasting events and find that, among ordinary people (nonexperts), "the correlation between price and overall rating is small and negative, suggesting that individuals on average enjoy more expensive wines slightly less." But experts might prefer more expensive wine; the study found that if wine A cost 10 times more than wine B, experts on average ranked it seven points higher on a 100-point scale. However, this effect was not quite statistically significant, and all that the authors can say with certainty is that experts don't *dislike* more expensive wine the same way normal people do.

Harrar et al.[3] have a study in *Flavour*, which was somehow a real journal until 2017, investigating novice and expert ratings of seven sparkling wines. Somewhat contrary to the point I made above, everyone (including experts) did poorly in identifying which wines were made of mostly red vs. white grapes (although most of the wines were mixed, which might make it a harder problem than just distinguishing pure reds from pure whites). More relevant to the current question, they didn't consistently prefer the most expensive champagne (£400) to the least expensive (£18).

Robert Hodgson[4] takes a slightly different approach and studies consistency among judges at wine competitions. If wine quality is real and identifiable, experts should be able to reliably judge identical samples of wine as identically good. In a series of studies, he shows they are okay at this. During competitions where wines are typically

1. Bénédicte Pineau et al, "Olfactory Specificity of Red- and Black-Berry Fruit Aromas in Red Wines and Contribution to the Red Bordeaux Wine Concept," *OENO One* 44, no. 1 (2010), https://doi.org/10.20870/oeno-one.2010.44.1.1457.

2. Robin Goldstein et al, "Do More Expensive Wines Taste Better? Evidence from a Large Sample of Blind Tastings," *Journal of Wine Economics* 3, no. 1 (2008): 1–9, https://doi.org/10.1017/S1931436100000523.

3. Vanessa Harrar et al, "Grape Expectations: How the Proportion of White Grape in Champagne Affects the Ratings of Experts and Social Drinkers in a Blind Tasting," *Flavour* 2, no. 1 (December 2013): 25, https://doi.org/10.1186/2044-7248-2-25.

4. Robert Hodgson, "An Examination of Judge Reliability at a Major U.S. Wine Competition," *Journal of Wine Economics* 3, no. 2 (2008): 105–13, doi:10.1017/S1931436100001152.

judged at between 80 and 100 points, blinded judges given the same wine twice rated on average about four points apart — in the language of wine tasting, the difference between "Silver–" and "Silver+". Only 10% of judges were "consistently consistent" within a medal range, i.e., they never (in four tries) gave a wine "Silver" on one tasting and "Bronze" or "Gold" the next. Another 10% of judges were extremely inconsistent, giving wine Gold during one tasting and Bronze (or worse) during another. Most of the time, they were just a bit off. Judges were most consistent at the bottom of the range — they always agreed terrible wines were terrible — and least consistent near the top.

In another study, Hodgson[5] looks at wines entered in at least three competitions. Of those that won Gold in one, 84% received no award (i.e., neither Gold, Silver, nor Bronze) in at least one other. "Thus, many wines that are viewed as extraordinarily good at some competitions are viewed as below average at others."

And here, too, a little bit of trickery can overwhelm whatever real stimuli people are getting. Lewis et al.[6] put wine in relabelled bottles, so that drinkers think a cheap wine is expensive or vice versa. They find that even people who had completed a course on wine tasting (so not quite "experts," but not exactly ordinary people either) gave judgments corresponding to the price and prestige of the labeled wine, not to the real wine inside the bottles.

So experienced tasters generally can't agree on which wines are better than others, or identify pricier wines as tasting better. Does this mean that wine is fake? Consider some taste we all understand very well, like pizza — not even fancy European pizza, just normal pizza that normal people like. I prefer Detroit pizza, tolerate New York pizza, and can't stand Chicago pizza. Your tastes might be the opposite. Does this mean there's no real difference between pizza types? Or that one of us is lying, or faking our love of pizza, or otherwise culpable?

5. Robert Hodgson, "An Analysis of the Concordance Among 13 U.S. Wine Competitions," *Journal of Wine Economics* 4, no. 1 (2009): 1–9, doi:10.1017/S1931436100000638.

6. Geoffrey Lewis et al, "The Impact of Setting on Wine Tasting Experiments: Do Blind Tastings Reflect the Real-Life Enjoyment of Wine?" *International Journal of Wine Business Research* 31, no. 4 (2019): 578–90, https://doi.org/10.1108/IJWBR-07-2018-0033.

I'll make one more confession — sometimes I prefer pizza from the greasy pizza joint down the street to pizza with exotic cheeses from a fancy Italian restaurant that costs twice as much. Does this mean the fancy Italian restaurant is a fraud? Or that the exotic cheeses don't really taste different from regular cheddar and mozzarella?

There can be objectively bad pizza — burnt, cold, mushy — but there isn't really any objective best pizza. Fancier and more complicated pizzas can be more expensive, not because they're better, but because they're more interesting. Maybe wine is the same way.

Notes on Notes

What about the tasting notes — the part where experts say a wine tastes like aged orange peel or avocado or whatever?

There aren't many studies that investigate this claim directly. But their claims make sense on a chemical level. Fermentation produces hundreds of different compounds, many are volatile (i.e., evaporate easily and can be smelled), and we naturally round chemicals off to other plants or foods that contain them.

When people say a wine has citrus notes, that might mean it has 9-carbon alcohols somewhere in its chemical soup. If they say chocolate, 5-carbon aldehydes; if mint, 5-carbon ketones.

(Do wines ever have 6-carbon carboxylic acids, or 10-carbon alkanes — i.e., goats, armpits or jet fuel? I am not a wine chemist and cannot answer this question. But one of the experts interviewed on *Somm* mentioned that a common tasting note is cat urine, but that in polite company you're supposed to refer to it by the code phrase "blackcurrant bud." Maybe one of those things wine experts say is code for "smells like a goat," I don't know.)

Scientists use gas chromatography to investigate these compounds in wine and sometimes understand them on quite a deep level. For example, from "Grape-Derived Fruity Volatile Thiols: Adjusting Sauvignon Blanc Aroma and Flavor Complexity":

> Three main volatile thiols are responsible for the tropical fruit nuances in wines. They are 3MH (3-mercaptohexan-1-ol), 3MHA (3-mercaptohexyl acetate) and 4MMP (4-mercapto-4-methylpentan-2-one). The smell is quite potent (or "punchy," as the Kiwis say) at higher concentrations, and descriptors used include tropical fruit,

> passionfruit, grapefruit, guava, gooseberry, box tree, tomato leaf and black currant. Perception thresholds for 4MMP, 3MH and 3MHA in model wine are 0.8 ng/L, 60 ng/L and 4.2 ng/L, respectively.

These numbers don't necessarily carry over to wines, where aromas exactly at the perception threshold might be overwhelmed by other flavors, but since some wines can have thousands or tens of thousands of nanograms per liter of these chemicals, it makes sense that some people can detect them. A few studies are able to observe this detection empirically. Prida and Chatonnet[7] found that experts rated wines with more furanic acid compounds as smelling oakier. And Tesfaye et al.[8] find good inter-rater reliability in expert tasting notes of wine vinegars.

Weil,[9] writing in the Journal of Wine Economics (another real journal!) finds that ordinary people can't match wines to descriptions of their tasting notes at a better-than-chance level. I think the best explanation of this discrepancy is that experts can consistently detect these notes, but ordinary people can't.

The Judgment of Paris

Until the 1970s, everyone knew French wines were the best in the world. Wine seller Steven Spurrier challenged the top French experts to a blind taste test of French vs. Californian wines. According to CNN:

> The finest French wines were up against upstarts from California. At the time, this didn't even seem like a fair contest — France made the world's best wines and Napa Valley was not yet on the map — so the result was believed to be obvious.
>
> Instead, the greatest underdog tale in wine history was about to unfold. Californian wines scored big with the judges and won in both the red and white categories, beating legendary chateaux and domaines from Bordeaux and Burgundy.
>
> The only journalist in attendance, George M. Taber of Time magazine, later wrote in his article that "the unthinkable happened," and in an allusion to Greek mythology called the event "The Judgment of Paris," and thus it would forever be known.

"The unthinkable" is, if anything, underselling it. One judge, horrified, demanded her scorecard back. The tasting turned California's Napa Valley from a nowhere backwater into one of the world's top wine regions.

7. Andrei Prida and Pascal Chatonnet, "Impact of Oak-Derived Compounds on the Olfactory Perception of Barrel-Aged Wines," *American Journal of Enology and Viticulture* 61, no. 3 (2010): 408.

8. Wendu Tesfaye et al, "Descriptive Sensory Analysis of Wine Vinegar: Tasting Procedure and Reliability of New Attributes," *Journal of Sensory Studies* 25, no. 2 (2010): 216–30, https://doi.org/10.1111/j.1745-459X.2009.00253.x.

9. Roman Weil, "Debunking Critics' Wine Words: Can Amateurs Distinguish the Smell of Asphalt from the Taste of Cherries?" *Journal of Wine Economics* 2, no 2 (2007): 136–44, doi:10.1017/S1931436100000390.

I bring this up because, well, the deliberately provocative title of this article was "Is Wine Fake?" Obviously wine is not fake: There is certainly a real drink made from fermented grapes. The real question at issue is whether wine expertise is fake. And that ties this question in with the general debate on the nature of expertise. There are many people who think many kinds of expertise are fake, and many other people pushing back against them; maybe wine is just one more front in this grander war.

And it would seem that wine expertise is real. With enough training (Master Sommelier candidates typically need 10 years of experience) people really can learn to identify wines by taste. Although ordinary people do not prefer more expensive to less expensive wine, some experts do, at least if we are willing to bend the statistical significance rules a little. And although ordinary people cannot agree on tasting notes, experts often can.

But although wine experts really do know more than you and I, the world of wine is insane. People spend thousands of dollars for fancy wine that they enjoy no more than $10 plonk from the corner store. Vintners obsess over wine contests that are probably mostly chance. False beliefs, like the superiority of French wine, get enshrined as unquestioned truths.

All the oenophiles and expert tasters of the 1960s and '70s got one of the most basic questions in their field wrong. Why? Maybe patriotism: Most of the wine industry was in France, and they didn't want to consider that other countries might be as good as they were. Maybe conformity: If nobody else was taking Californian wines seriously, why should you? Or maybe a self-perpetuating cycle, where if any expert had made a deep study of Californian wines, they would have been able to realize they were very good, but nobody thought such a study was worth it.

Wine is not fake. Wine experts aren't fake either, but they believe some strange things, are far from infallible, and need challenges and blinded trials to be kept honest. How far beyond wine you want to apply this is left as an exercise for the reader.

76

China's Silicon Future

Karson Elmgren

China dreams of competing with global superpowers in the semiconductor industry. Whether its efforts will succeed is far from clear.

ILLUSTRATIONS BY
Gizem Vural

In 2022, American politicians cannot agree on much. One of the few things they do see eye to eye on is the importance to U.S. national interests of the semiconductor industry, which undergirds technological progress in everything from the humble household robotic vacuum to the most sophisticated military hardware. This consensus has resulted in one of the crowning legislative achievements of Joe Biden's presidency so far: the CHIPS and Science Act. The law allocates over $50 billion to support the U.S. semiconductor industry, even though the United States already dominates entire industry sectors. Why is America so anxious about chips? The answer is that mastering semiconductor manufacturing has proven both an economic engine and an indispensable strategic asset for the states that have managed to do so. The few who have gained such a position are determined to keep it; the rest covet a spot among their ranks. China, in particular, has made growing a well-rounded domestic industry a national priority, motivated in large part by its techno-economic grappling match with the United States. Whether its efforts will succeed remains to be seen.

Semiconductors — chips — are the foundational components that underlie almost all of the modern electronics industry. They are what allow computers to compute — and in today's world, almost everything is a computer. Conventional semiconductors are made from silicon, typically refined from sand, and produced by one of the most sophisticated, laborious and globalized value chains of any industry. It is traditionally segmented into three steps: design, manufacturing and packaging. These three steps in turn rely on fancy gadgets galore, built, operated and serviced by initiates of technical arts from software engineering to materials science, and raw materials ranging from highly pure silicon to xenon gas.[1] Designing a chip requires arranging billions of components on a postmark-sized bit of metal and wrestling against quantum mechanics to get them to work together properly.

1. The absolute minimum purity needed for electronics-grade silicon is 99.9999999%, or one nonsilicon atom to every billion silicon atoms. Some producers offer silicon that is a full two orders of magnitude purer still, with a single nonsilicon atom to every 100 billion silicon atoms.

All those billions of components must be carved out of silicon with lasers accurate enough to hit a pingpong ball on the moon. Chip factories, referred to as foundries or "fabs," are some of the most fastidiously controlled environments on the planet. The precautions taken to avoid letting the faintest trace of contamination into fabs are rivaled only by those taken to avoid letting pathogens out of high-security biolabs. Staff must wear full protective equipment. Moisture is removed from the air that enters the facility to filter errant particulates; the air is then re-humidified to optimal levels for manufacturing. Coordinating the resources and talent to produce chips at scale is a subtle art not easily mastered or distilled to a repeatable method.

Conveniently for American policymakers, the semiconductor industry happens to be almost entirely dominated by the United States and its close partners and allies in Japan, South Korea, Taiwan and Europe. Most semiconductor design is done by American firms. All three companies offering the specialized software tools used for the process are based in the U.S. Taiwan Semiconductor Manufacturing Company, or TSMC, is responsible for over 90% of manufacturing of the most advanced chips; South Korea makes the rest. The machinery required to make those cutting-edge chips is produced by only one company: the Dutch firm ASML. Though it has lost a leading position in other parts of the industry, Japan today almost entirely controls photoresist processing, a key step in manufacturing, and plays an important role in photoresist materials as well.

For China, depending on its main competitors for critical components is unacceptable. The Chinese Communist Party knows a thing or two about economic leverage, having itself used its dominant position in the processing of rare earth metals to threaten or punish other countries since 2010.[2] As tensions have risen in recent years, the United States has increasingly moved to restrict China's access to semiconductor technologies, challenging the CCP's dreams of technological empire. Beijing's "no-limits" partner to the north, Russia, has tasted firsthand what it feels like to have semiconductor imports decimated by American sanctions. Allegedly, Russian military equipment contains chips salvaged from dishwashers and refrigerators, while Yale researchers describe the outlook for Russia as "economic oblivion." China's dependence on the semiconductor supply chain likely moderates the country's willingness to escalate international conflicts with the United States and its partners.

Conversely, if China were to achieve an entirely domestic supply chain for all its semiconductor needs, it may be emboldened to throw its weight around on the world stage. Taking the lead in chips would generate leverage that may make other countries even more hesitant to anger Beijing. In short, the business of making sand do math may be the most important

2. Chinese state-affiliated hacking groups have also recently attempted to undermine American attempts to build a domestic rare earths supply chain through a disinformation campaign against a planned rare earths mining facility in Texas.

civilian industry for the future of China's "comprehensive national power," to put it in CCP jargon.

***

China's silicon ambitions — and struggles — are not new. By the 1970s, Mao Zedong was gone and the tumult of the Cultural Revolution had ended, but the country was in shambles. In 1978, Deng Xiaoping's top priority as newly instated leader of the CCP was development, and development required technology. China launched several state-led semiconductor projects between the mid-1980s and the mid-1990s, but they were hampered by administrative delays, while international progress forged ahead. When the IT bubble popped in 2001, some international firms cut their losses in the Chinese market, leaving their partners to wither away. Piecemeal investment from domestic firms and government created factories that were outdated almost as soon as they got up and running. Perhaps worst of all, the generation of engineers who should have led the industry in the '80s and '90s had spent their formative years doing hard labor instead of training in their field — if they had survived the Red Guards' persecution at all. By the turn of the 20th century, China was left with little to show for its efforts.

With these projects, CCP leaders had been hoping to follow a script that had proven successful for its neighbors. Japan, having clawed its way into the chip industry through state-led industrial policy during its postwar boom, was a model.[3] By the 1980s, American chipmakers were bemoaning the loss of the memory market to Japanese shops whose process innovations had cut defect rates, minimizing waste and boosting profitability. Other of East Asia's burgeoning economic powerhouses had begun to get in on the action as well. In South Korea, support from both government and business conglomerates helped the semiconductor industry grow at breakneck speed in the mid-1990s. In 1987, Morris Chang (張忠謀 *Zhāng Zhōngmóu*), then head of Taiwan's Industrial Technology Research Institute, founded Taiwan Semiconductor Manufacturing Company with 21% of the company's startup capital provided by the Taiwanese government. But besides avoiding the autolobotomous chaos of the Cultural Revolution, these countries had another advantage China lacked: decades of commercial and intellectual trade with the United States, where many engineers had been trained at cutting-edge academic and industrial labs before returning to build suppliers, customers and competitors to their former employers in their home countries. Tacit knowledge developed through hands-on experience has been crucial for success in the semiconductor industry.

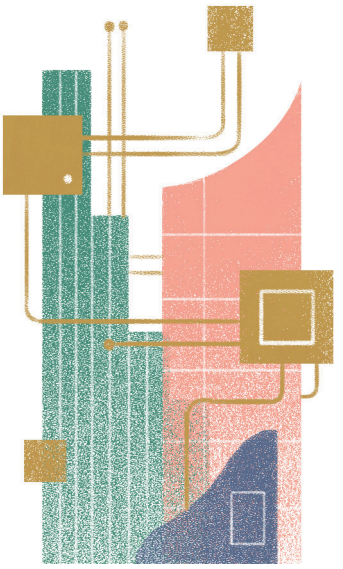
Over 20 years, Deng's policies gradually replenished China's pool of technical talent. Reform and opening revived China's educational institutions such as Tsinghua University, which

3. Japan's Ministry of International Trade and Industry, which directed much of Japanese industrial policy through the country's economic miracle, earned the epithet "the notorious MITI" from foreign commentators and inspired the term "Japan, Inc." due to its powerful influence in Japan's economy.

would become a key crucible of Chinese semiconductor talent. It also meant that Chinese students began to flow into classrooms at places like MIT and Berkeley, and later into the labs where alumni of such schools often end up. As China's economic growth took off in the early 2000s after the country was admitted to the World Trade Organization, many Chinese educated abroad took a chance and returned to China, forming a great migration of "sea turtles."[4] This new generation of world-wise leaders had the know-how to establish themselves in the industry; some founded firms that would later become major players. However, the government support that had midwifed semiconductor industries around the world was not forthcoming. Through the first decade of the new millennium, bureaucrats had a bitter taste in their mouth from the failures of large-scale funding to achieve results in attempts of the previous decades.

***

All this changed when Xi Jinping came to power in 2013. Unlike his predecessors, who experimented with free market reforms, Xi has moved to reassert strong central control over the Chinese economy. Most notably, this has taken the form of the now-infamous Made in China 2025 industrial strategy, a trillion-dollar initiative to upgrade China's high-tech manufacturing capacity. Xi has also spoken at length about the importance of the "real economy" (实体经济 *shítǐ jīngjì*) — solid, utilitarian sectors like manufacturing



and infrastructure rather than social media, financial speculation or video games. It is no surprise that the CCP's enthusiasm for industrial policy came to semiconductors. In 2014, the Chinese government laid out an ambitious set of goals to develop a world-leading semiconductor industry by 2030. The same year, the government established the China Integrated Circuit Industry Investment Fund (国家集成电路产业投资基金 *Guójiā Jíchéng Diànlù Chǎnyè Tóuzī Jījīn*), more often referred to simply as the "Big Fund," targeting $19 billion of direct investment over five years. Semiconductors are both the most critical sector of the real economy for China's resilience and the apotheosis of Xi's economic aesthetics: high technology, tangible objects and strategic leverage.

4. 海龟 *hǎiguī*, literally "sea turtle," is a pun on the phrase "overseas returnee," 海归, also pronounced *hǎiguī*.

Policy support under Xi has focused on remedying China's weaknesses in the technically demanding, high-margin sectors of design and manufacturing. One of the country's greatest successes is its analog of TSMC: Semiconductor Manufacturing International Corporation. TSMC disrupted the industry in the 1990s with the "pure-play foundry model," focusing exclusively on manufacturing over design. SMIC is another pure-play foundry. Like TSMC, it was founded by an American-educated returnee with a decades-long career at the U.S. semiconductor firm Texas Instruments. In July 2022, reverse-engineering firm TechInsights reported that a chip produced by SMIC had been made with technology just one generation behind the commercial state of the art — a significant accomplishment even if not yet indicative of a full commercial-scale production capability.[5] As semiconductor technology has advanced, the ranks of foundries capable of producing the smallest, most sophisticated chips have thinned from dozens in the late 1990s to a mere trio today: TSMC, Samsung and Intel. Although SMIC will need to run even faster to catch up with, much less overtake, the leaders, maintaining a fast-follower position is impressive in its own right.

Although American firms such as NVIDIA and Intel dominate design, China has managed to cultivate a number of firms capable of dreaming up advanced chips. HiSilicon, a subsidiary of Chinese telecoms giant Huawei, produced logic chips on TSMC's 5-nanometer production node before falling under U.S. sanctions.[6] A swath of Chinese firms aim to challenge NVIDIA's vise grip on GPUs, with some producing designs only a few generations behind NVIDIA's. In the somewhat less challenging memory market, China has produced strong players such as Yangtze Memory Technologies Corp, which grew from 1% of global market share at the start of 2020 to 5% as of summer 2022, and secured a deal to supply memory chips for the newest iPhone.

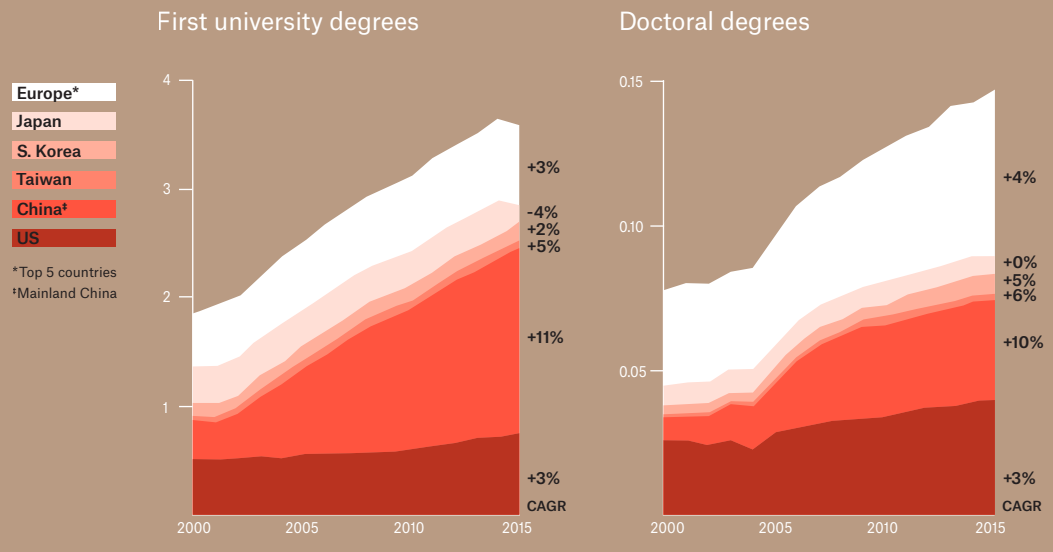
As semiconductor technology has advanced, the ranks of foundries capable of producing the smallest, most sophisticated chips have thinned from dozens in the late 1990s to a mere trio today: TSMC, Samsung and Intel.

China is strongest in the less sexy but no less critical final link of the semiconductor value chain: assembly, testing

5. The chip was reportedly "a near duplicate of TSMC 7nm process technology," leading some to speculate that industrial espionage may have played a role in the achievement.

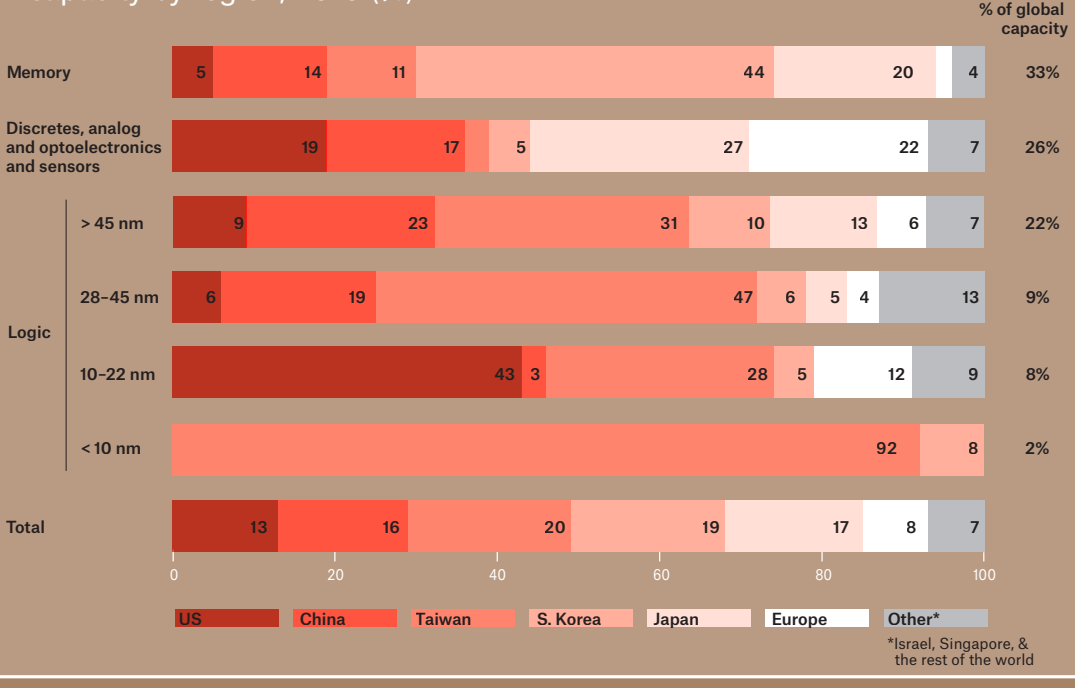
6. UNISOC has since become the leading Chinese designer of smartphone processors, with products on TSMC's 6-nanometer node.

Annual graduates in science and engineering (millions)



Source: BCG analysis with data from US National Center of Science and Engineering Statistics (NCSES)
https://www.semiconductors.org/wp-content/uploads/2021/05/BCG-x-SIA-Strengthening-the-Global-Semiconductor-Value-Chain-April-2021_1.pdf

Breakdown of global wafer fabrication capacity by region, 2019 (%)



Source: https://www.whitehouse.gov/wp-content/uploads/2021/06/100-day-supply-chain-review-report.pdf

and packaging (abbreviated collectively as ATP). ATP generally requires a lower level of skills, more akin to the low-end manufacturing and device assembly work that has been China's bread and butter for decades. It also garners lower margins. Nevertheless, it is a critical sector in which China made up almost 40% of the global market as of 2021.

Furthermore, advanced packaging is an area of active innovation in the industry where China may be well-positioned to lead if it is so inclined. Though it has to date employed primarily low-skill labor, greater complication and precision required by fancy packaging methods such as three-dimensional stacking means the industry will likely move toward heavier use of industrial robots. As the largest consumer of industrial robots for almost a decade running, China could potentially parlay its robotic manufacturing capacities into a significant moat.

***

Chip mania reached a fever pitch throughout the Chinese economy by 2019, following a set of new tariffs introduced by the Trump administration and the subsequent realization that U.S.-China trade tensions would neither go away nor be quickly and easily resolved through a trade deal. The second tranche of investment from the Big Fund was announced in October of that year, promising another $20 billion focused on equipment and upstream materials. Tellingly, Xi also placed top economics official and close confidant Liu He, previously responsible for China's end of the U.S.-China trade war, in charge of semiconductors policy in June 2021. Giving semiconductor manufacturing a high position among policy priorities also means the party is more willing to pay the political costs of encouraging industrial espionage, protectionist policies and other measures that may benefit the industry.[7]

As is characteristic in China's highly policy-driven market, private investment has stampeded to position itself in this "wind tunnel" of opportunity.[8] According to Chinese state sources, over 58,000 new chip firms were registered in China between January and October 2020 alone. Some firms signal their red bona fides with patriotic marketing decisions: one advertises a "Self-Strengthening" line, which uses only Chinese domestic components. Nearly every major Chinese tech company has made an foray into the sector, no matter how far their core consumer-facing software business might be from the primarily B2B computer hardware market. Even TikTok's parent company, ByteDance, has thrown its hat into the ring. In some cases, designing hardware that can be co-optimized with the rest of a company's technology stack can provide real benefits; in others, these firms may be hoping to cash in on government incentives, or simply signaling their piety in following the CCP's leadership.

China has a few natural advantages to give all these new companies a head start. As in many domains, one is its

7. China is not unique in using tactics that violate the rules and norms of international trade to build its chip industry; Japan did many of the same things in the 1980s as it strove for semiconductor success.

8. *Fēngkǒu*, literally a gap or hole through which wind flows, is a term often used in the Chinese business world to refer to promising sectors where favorable economic "winds," often policy-generated, will blow in the sails of hopeful firms.

sheer size. To remain in the lead, chip companies must spend massively on R&D, shoveling up to 15%-30% of revenue back into the engine of technological progress.[9] China's uniquely large domestic market means firms can establish a flywheel of sales to R&D to more sales without needing to venture into other markets. In recent years, China has been the world's second-largest consumer market for semiconductors, comprising a quarter of global demand in 2019, about on par with the United States. In addition to demand, China has a major strength in supply — of labor, at least. China is projected to graduate over 77,000 STEM Ph.D.s in 2025, compared to only 40,000 in the United States. Chinese universities have established a swath of semiconductor-related degree programs over the last years to train more specialists in the arcane arts of building thinking machines.

But while the inputs to the industry driven by state support are impressive, it is far from guaranteed that their outputs will be equally so. Mandarin Chinese offers a pun to encapsulate a major issue for the industry: The second character in the word for "chips," *xīnpiàn*, can be swapped for a homophone 骗 *piàn* meaning "to swindle," creating a neologism 芯骗 *xīnpiàn*, "chip cheating." Of the legions of "semiconductor" firms registered in recent years, it is unclear not only how many will succeed, but how many are even contributing to the industry in any meaningful sense. The story of one of China's older companies, Wuhan Hongxin Semiconductor Manufacturing Company, provides a case in point.[10] Sometimes referred to as China's "semiconductor Theranos," HSMC was a scam led by a career con man and two accomplices with no technology expertise who succeeded in roping in the founding CTO of TSMC, acquiring a top-of-the-line lithography machine, and raising billions of dollars before plundering the firm's assets and making off scot-free. Even the Big Fund itself has been beset by graft scandals, with at least four current or former senior executives involved in managing the fund coming under investigation since July 2022.[11]

But while the inputs to the industry driven by state support are impressive, it is far from guaranteed that their outputs will be equally so.

American policymakers, recognizing the geopolitical strategic importance of semiconductors, have started to take

9. By comparison, R&D spending as a fraction of revenue in sectors such as energy or consumer products can be as low as 1% or even 0.1%.

10. Also abbreviated as HSMC, the company's name 弘芯 *Hóngxīn*, "grand chips" or "promoting chips," patriotically echoes the term "red heart," 红心 *hóngxīn*, used to describe passion, especially for the Communist Party.

11. This includes Ding Wenwu, Wang Wenzhong, Lu Jun and Yang Zhengfan.

action to stymie China's efforts. They have levied or threatened sanctions on multiple Chinese semiconductor players, including Huawei, ZTE and YMTC. They have also made use of the choke points within the industry. Dutch firm ASML is the sole manufacturer of the extreme ultraviolet lithography equipment necessary to make the most advanced chips, and it has already agreed to American entreaties not to sell its most advanced kit to China. In recent months, Washington has introduced a bevy of new measures to hamstring China's semiconductor industry, including restricting exports of even older semiconductor manufacturing equipment, banning American companies from servicing equipment in China, requiring export licenses to ship top-shelf GPUs to China and planning to block certain Chinese firms and research labs from purchasing anything made with a drop of US technology. These new policies are designed to restrict the progress of China's industry past the generation of technology it has already mastered, locking the country out of high-end chips for applications like AI and supercomputing entirely.[12]

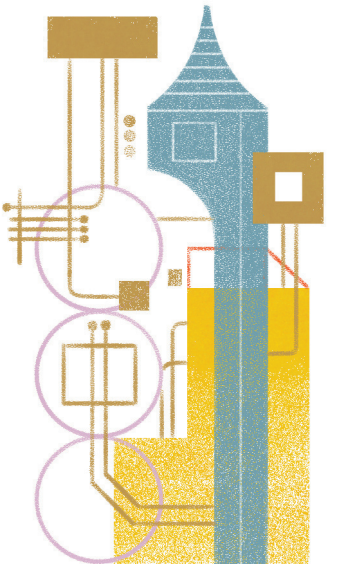
Beyond Beijing, a similar chip mania has infected governments around the world whose ambitions in the space further threaten to crowd out China's place in the silicon sun. In the United States, the CHIPS Act has finally passed, with measures to bolster the United States' strategic position in the industry, particularly in leading-edge fabrication. Taiwan, South Korea, Japan and Europe have all explored new policy support for the industry as well. The leading firms have geared up for the melee too. TSMC announced in January 2022 it would spend $100 billion over three years. Samsung funneled over $40 billion into semiconductors in 2021 alone. Meanwhile, the United States continues to dominate in R&D, making up more than half of total industry spending. What was already a competitive market may become even more cutthroat. China will not find itself uncontested in chasing the various prizes that await those who solve the engineering puzzles holding back progress today.

Flows of technical knowledge into China are also being constricted. As University of Toronto professor Jun Zhang put it in a recent talk: "To overtake the leaders, you have to learn from them. But will they give you opportunities to learn from them?" China's success to date has been catalyzed by knowledge and experience gained from foreign industry leaders. They are no longer so generous in offering such education. Taiwan's government has introduced anti-poaching regulations to cut down on the trend of Taiwanese semiconductor engineers being lured to the People's Republic, where salaries may be multiples of their compensation on the island. Investment flows from China into global semiconductor firms have met with increasing resistance since the Trump administration blocked a Chinese private equity firm from acquiring Lattice Semiconductor in 2017. Now even investments into China are garnering closer scrutiny as avenues for leakage of technical know-how. All but the most oblivious

12. Prospects are better for older generations of semiconductors, which have not been subject to the same restrictions and are still crucially important for a panoply of simpler electronics.

international semiconductor firms recognize they are targets of Chinese state-affiliated hacking groups and are actively hardening their defenses against espionage.

If China's plans are to succeed, they will do so through the assiduous efforts of a motivated, masterful semiconductor workforce. But so far, "hard tech" sectors have struggled to attract talent. The Chinese tech industry is infamous for its grueling conditions, often summarized as "996" — 9 a.m. to 9 p.m., 6 days a week. An entire lexicon has emerged to describe the crushingly competitive professional lives of urban white-collar workers. The dominant concept is "involution," (内卷 *nèijuǎn*), a race of ever-increasing competition over finite resources that leaves all participants burnt-out and despondent. "Lying flat" (躺平 *tǎngpíng*) has become a rallying cry of slackers. The working masses in China sometimes refer to themselves as "chives" (韭菜 *jiǔcài*), growing tall only to be harvested as cheap labor by the combine of state-led capitalism. The semiconductor industry, despite its pride of place in the CCP's eye, is far from exempt from these dynamics. The training in electrical engineering necessary to contribute in the industry — generally considered to require a Ph.D. and significant time gaining hands-on experience — is a heavier burden than the comparatively relaxed investment of time needed to break into software, where salaries are similarly high and no one needs to spend their days wrapped up in PPE in a clean room.[13] Why, some might



wonder, go through all the effort if you will only be chopped up with the rest of the chives?

Optimistically, China's flood of semiconductor investment could cultivate a pool of human capital that will form the foundation for the industry's eventual success. However, the country may equally just waste the time of technical talents who could otherwise be pursuing more economically efficient, market-driven opportunities. In flushing the intensely competitive and challenging manufacturing sector full of cash, the Big Fund's Phase I investments may be fueling involution in chipmaking while failing to capitalize on opportunities for innovation elsewhere, such as in advanced packaging.

13. The same dynamic also affects the semiconductor industry elsewhere; it is not uncommon for even newly minted computer hardware Ph.D.s to immediately jump ship for the milder climes of software.

Meanwhile, the Chinese economic engine is becoming creakier as the population ages and the debt-fueled infrastructure investment that powered growth for decades runs into diminishing returns. Official state estimates forecast population growth to peak in 2031, but some researchers believe the birth rate has already begun to decline, and may in fact have peaked as early as 1991. Concerning signs have surfaced in property markets, a critical part of China's financial stability. Strict zero-COVID policies pose a dilemma for the CCP: They incur high economic and public opinion costs, but lifting them risks an outbreak, which could itself also cause economic havoc.[14] Xi Jinping, whose personal influence over Chinese policy is perhaps greater than any leader since Mao Zedong, lacks a strong background in economics and is suspected to make decisions based more on his expertise in domestic politics and foreign policy. Though the People's Republic has many competent technocrats, Xi's word may override the wisest counsel.

Some think he already has with the new wide-ranging package of economic and social reforms dubbed the "Red New Deal." The hammer has come down particularly hard on Chinese consumer tech companies, which have been subject to anti-monopoly measures and other new restrictive regulations. While some of these policies appear commonsensically technocratic, the party's recent appetite for bold action has shown itself most clearly in policies related to children, including banning after-school tutoring services (previously a boisterous market) from turning a profit, and limiting minors to only three hours of video game play per week — Friday, Saturday and Sunday from 8 p.m. to 9 p.m. While intended to promote "common prosperity," these policies may end up hamstringing the private sector dynamism responsible for almost all

Success in semiconductors would be not only a source of national pride for the Chinese Communist Party, but a major shift in the global distribution of techno-economic power.

14. China has to date refused to accept the more-effective American and European mRNA vaccines and has a low rate of vaccination even with its less-effective domestic vaccines. Health care capacity is also highly unequally distributed in the country, meaning the system could easily become overloaded in rural areas.

15. In particular, cutting off exports of chips, rather than the tools and materials to make them, opens up space in the domestic market for Chinese chip firms that would previously have been dominated by the NVIDIAs and AMDs of the world, helping them invest in the R&D necessary to get to the cutting edge. The U.S. recently imposed export license requirements on top-shelf GPUs, which could have exactly this effect.

16. The aforementioned field of advanced packaging, which includes methods such as simply stacking chips on top of one another, is one way the field may yet continue to squeeze more performance out of conventional silicon.

China's economic growth. The Chinese government in late August announced a large stimulus package, but even this was not enough to buoy economists' and investors' hopes for the Chinese economy. The future is far from certain for the Chinese economic juggernaut.

***

What, then, should the savvy observer expect from China's chip dream? The country hoped to source 40% of semiconductors domestically in 2020. Ultimately, it reached 16%. We might conclude that today's campaign is as doomed as in the past. The semiconductor industry today requires even more exacting technical sophistication. But until quite recently China remained poor; today's China is a vastly more formidable technological competitor. American efforts to cut off China's semiconductor industry from global supply chains may even backfire and help it to grow strong and independent.[15] Most likely, if trends continue, the country will achieve a middling outcome — making a respectable showing in older technology, perhaps eking out enough progress to inch forward a few generations, and likely achieving a handful of major technical breakthroughs.

Yet, there is one other dynamic that may provide China with surprising opportunities. Moore's law, the self-fulfilling prophecy of predictable progress in shrinking transistors to make chips faster and more efficient, has guided the industry for over half a century. It is now coming to an end as the conventional technical paradigm runs into fundamental physical limits.[16] Scientists have proposed and prototyped alternative architectures for computing hardware, each with its own advantages and constraints. Many industry observers expect the future of computing to see heterogeneous components stitched together in tailored combinations to suit specific cases, rather than the handful of general-purpose architectures ubiquitous today. We should expect that China may succeed in at least a few of these directions. The resulting strategic landscape may look very different depending on which work out and which do not. If, for instance, China pulls ahead in quantum computing, where it is relatively strong, but cannot produce ultrawide bandgap semiconductors, recently placed under export control by the U.S. Commerce Department, Beijing may find itself able to crack the toughest encryption in the world but struggling to field satellites with spyware to match those of its competitors. If the United States and its partners continue to tighten the screws, the number of such successes may be very limited. On the other hand, even one major success could give China a powerful position if every international firm is clamoring to use, for example, Chinese packaging processes, much as TSMC commands great influence through its near monopoly on the most advanced production methods.

Success in semiconductors would be not only a source of national pride for the Chinese Communist Party, but a major shift in the global distribution of techno-economic power. It would be positive evidence of a strength of the Chinese model and a reinforcing buttress for it in the future. Failure, on the other hand, could lead to stagnation.

90

The Illogic of Nuclear Escalation

Fred Kaplan

How much is enough? It's the most basic question in the nuclear arms race. For over sixty years, few have asked it, and even fewer have received an answer.

This past summer, a bipartisan majority of Congress, with the blessing of President Biden, approved a massive military-spending bill that included $51 billion for nuclear weapons — nearly 20 percent more than allotted by the previous year's budget, which itself broke previous records. Earlier in the spring, the Biden administration sent to congress a Nuclear Posture Review, committing to upgrade all three "legs" of the "strategic Triad" — including a new missile-launching submarine, a new bomber and a new land-based intercontinental ballistic missile — as well as a bevy of new bombs and warheads for these weapons to launch or drop. Since these weapons are still in development or the early phases of production, the costs are bound to grow; the price tag for the refurbished Triad alone is estimated at $2 trillion over the next 30 years.

The official rationale for this upgrade is that the existing subs, bombers and ICBMs are approaching obsolescence. Even if this claim were true (more about that later), it begs the question of whether the arsenal needs to be as large as it is. A serious assessment of the arsenal must begin by asking "How much is enough?" and, its corollary query, "Enough to do what?"

Yet in the debate over America's nuclear stockpile, to the extent there is debate, these questions are going unasked. It is hard to have an informed *public* debate, as many of the issues are classified, esoteric or both. But even the debates in Congress and inside the executive branch tend to be shallow. Almost nobody is asking those basic questions. In fact, in the 60-plus years of the nuclear arms race, almost nobody ever has.

It's important to examine the secret history of this race to understand not only how we got here but how to ask those questions — and how to change course.

The Race Is On

Momentum drove the nuclear arms buildup from the beginning. There never was a decision to drop the atomic bombs on Hiroshima and Nagasaki at the end of World War II. Once it was decided to build the bombs two years earlier, it was inevitable they would be dropped.[1] The United States had just two A-bombs in early August 1945; the Japanese emperor surrendered before a third bomb was ready to go.

After the war, as the Cold War got underway, it was assumed — no doubt correctly — that the Soviets would build A-bombs once they figured out how (they exploded their first roughly four years after Hiroshima); so the American weapons labs preemptively churned out more bombs. Once American scientists tested the much more powerful hydrogen bomb in 1952, there was little doubt that it too would be built (though many who had helped build the A-bomb protested) or that the Soviets would build their own (which they did three years later).

President Dwight Eisenhower, a retired five-star Army general and WWII commander, was not at all bloodthirsty; once he understood the power of nuclear weapons, he feared and detested them. But he also believed, as did most officials and analysts, that if the U.S. and the USSR ever locked arms, even in a "small" war over a narrow strip of territory in Europe or Asia, it would soon escalate to nuclear exchanges. So the wise policy would be to deter the Soviets from attacking in the first place, and the best way to do that, he figured, was to warn them that we'd blow them to smithereens if they did. His secretary of state, John Foster Dulles, called the policy "massive retaliation," and the Joint Chiefs of Staff — composed of the top U.S. military officers — translated it into a war plan that italicized *massive*.[2]

Few realized at the time, or in the years since, just how massive it was. By 1960, the U.S. war plan called for launching the entire nuclear arsenal — at the time, 3,423 weapons, exploding with the blast power of 7,847 megatons — against 1,043 targets in the Soviet Union, its satellite countries in Eastern Europe and Communist China.[3] This was not a plan to strike back if the Soviets launched a nuclear attack on the U.S.; it was a plan to strike *first* if the Soviets mounted a *non-nuclear* invasion against U.S. allies.

Some in Washington asked how many people such an attack would kill. The answer that came back from those who devised the war plan at Strategic Air Command (SAC) in Omaha was *275 million*.[4] Such a figure had previously been inconceivable. No one could imagine a war aim that required killing so many civilians.

1. Barton J. Bernstein, "The Atomic Bombings Reconsidered," *Foreign Affairs*, Jan. 1995.

2. Eisenhower Diary Series (Jan. 1956 Diary), "Net Evaluation of Damage Anticipated in Initial Stages of US-USSR Nuclear War," Box 12, Ann Whitman File, Eisenhower Library; John Foster Dulles, "The Strategy of Massive Retaliation," Council on Foreign Relations, Jan. 12, 1954; JCS 2101/244, "Strategic Concept for General War," Mar. 29, 1954, National Archives/Modern Military Branch.

3. Memo, Op-06 to Op-00 on the Initial NSTL & SIOP, Nov. 22, 1960, Arleigh Burke Papers, Memos & Letters (NSTL), Navy Yard.

4. *Wizards of Armageddon*, 269; Daniel Ellsberg, *The Doomsday Machine* (Bloomsbury, 2017).

What's striking is that, even so, no one among the few officials privy to this plan questioned its validity or how the numbers were calculated. They never asked whether such a massive arsenal, or such a cataclysmic attack, was necessary for national security.

The plan was founded, in large part, on the basis of self-interest. SAC — the branch of the now-independent U.S. Air Force that controlled nuclear plans and operations — had set up a unit called the Joint Strategic Target Planning Staff (JSTPS). Its job was to find every plausible target inside the Soviet empire, then assign U.S. nuclear weapons to destroy each one. As JSTPS found more targets, SAC had a rationale to request more weapons. As the Soviets matched the U.S. arms buildup, they created more targets — thus driving the rationale for still more U.S. bombs and warheads.

As an added twist, JSTPS declared that some of these targets were deemed more "high-value" than others. At the time, no single nuclear bomb had as much as an 80 percent chance of destroying a particular target on the other side of the world, partly because they were inaccurate, partly because some might be duds. So more than one weapon — in some cases three or even more weapons — would have to be aimed at each of those targets in order to achieve the required level of destruction. In the first formal war plan in 1960, known as the SIOP (for Single Integrated Operational Plan), JSTPS designated targets and specified for each a minimum probability that it would be destroyed. Seven of the most important, hard-to-hit targets needed to be targeted with a probability of 97 percent, 213 targets with 95 percent, 592 targets with 90 percent and 715 targets with 80 percent.[5] Raise the required level of destruction — that raised the required number of weapons.

All of this was more highly classified than anything else in government. These decisions and calculations — which crucially affected how many weapons the government would build — were made with no input from any officials, or even military officers, in Washington. No part of the plan was declassified until decades later.[6] Still, through the decades, the Pentagon and Congress routinely approved SAC's requests for more weapons.

The Logic of Mutually Assured Destruction

As this self-propelled nuclear arms race hurtled into motion, two obstacles reared their heads. First, as the Soviets started producing and deploying their own nuclear weapons, "massive retaliation" became a suicide pact. If the Soviets invaded West Germany and the president clobbered Moscow and other Russian cities with nukes, the Kremlin would clobber New York and other American cities right back.

The second obstacle came in the form of a bureaucratic battle inside the Pentagon. At the time, nuclear weapons dominated the military budget (much of the Army, which fought World War II and the Korean War,

5. Messages, CINCLANT to CNO, Nov. 22, 1960, Apr. 27, 1961, Arleigh Burke Papers, NSTL Messages, Exclusives & Personals.

6. I got the memos declassified for *Wizards of Armageddon*. Some of them were subsequently reclassified, then many years later declassified again.

had been demobilized). The Air Force controlled the arsenal, and so dominated the budget. But in the late 1950s, the Navy, whose battleships had been eclipsed by the Air Force in the budget wars, designed and built the Polaris submarine — a vessel that could carry 16 nuclear missiles and launch them while moving, undetected, underwater. Analysts had observed that Air Force bombers and land-based missiles would soon be vulnerable to a Soviet first strike; the Air Force cited this claim as another excuse to request funds for *more* bombers and missiles. But analysts in the Navy's official think tank came up with a new strategic idea: *finite deterrence*. All the U.S. needed, they argued, was enough nuclear weapons to destroy the 100 largest cities in the Soviet Union; this could be done with a mere 640 missiles in 40 submarines.[7] If the Soviets built more weapons, it wouldn't matter; as long as the U.S. had the submarines, which the Soviets couldn't track, it could destroy those cities and deter the Soviets from starting a war to begin with. Unlike the Air Force, the Navy could ensure deterrence without setting off an arms race.

Air Force generals were panicked, so they turned to their own think tank, the RAND Corporation, where analysts had devised a counter-idea: If the Soviets invaded Western Europe, the U.S. should respond not by firing all its nukes at once but by launching a *limited* nuclear attack only against the Soviets' ICBMs, bomber bases and submarine ports. They could then tell the Kremlin, "If you continue your aggression, we will use our remaining arsenal to destroy your cities." The idea was to stop the war before it careened out of control.

At first, the idea did not appeal to the Air Force generals: The whole point of a nuclear bomb, to them, was that it inflicted massive damage; they had no interest in war plans that put the arsenal's power on a leash. But the RAND idea required bombs that were accurate enough to hit an airstrip or a missile base without damaging a nearby city. Bombs dropped from Air Force bombers could do that; Polaris missiles, which were notoriously inaccurate, could not. So the generals adopted the strategy, known as "counterforce" — at least rhetorically — as a way of beating back the Navy's challenge. (The RAND memorandum outlining the strategy was titled "The Polaris Problem.")[8]

Robert McNamara, secretary of defense under President John F. Kennedy, also liked the idea, but for different reasons. Days after he entered the Pentagon, he was briefed on SAC's nuclear war plan — which would launch the entire U.S. arsenal, as quickly as possible, at every target in the Communist world — and was appalled by its inhumanity. He and his aides ordered SAC to revise the plan to give the president "options" to launch various limited attacks.[9]

Within a few years, though, McNamara discovered a downside to his order. If the plan entailed destroying all of the Soviets' nuclear weapons, then for every new nuke that the Soviets deployed, the U.S. would

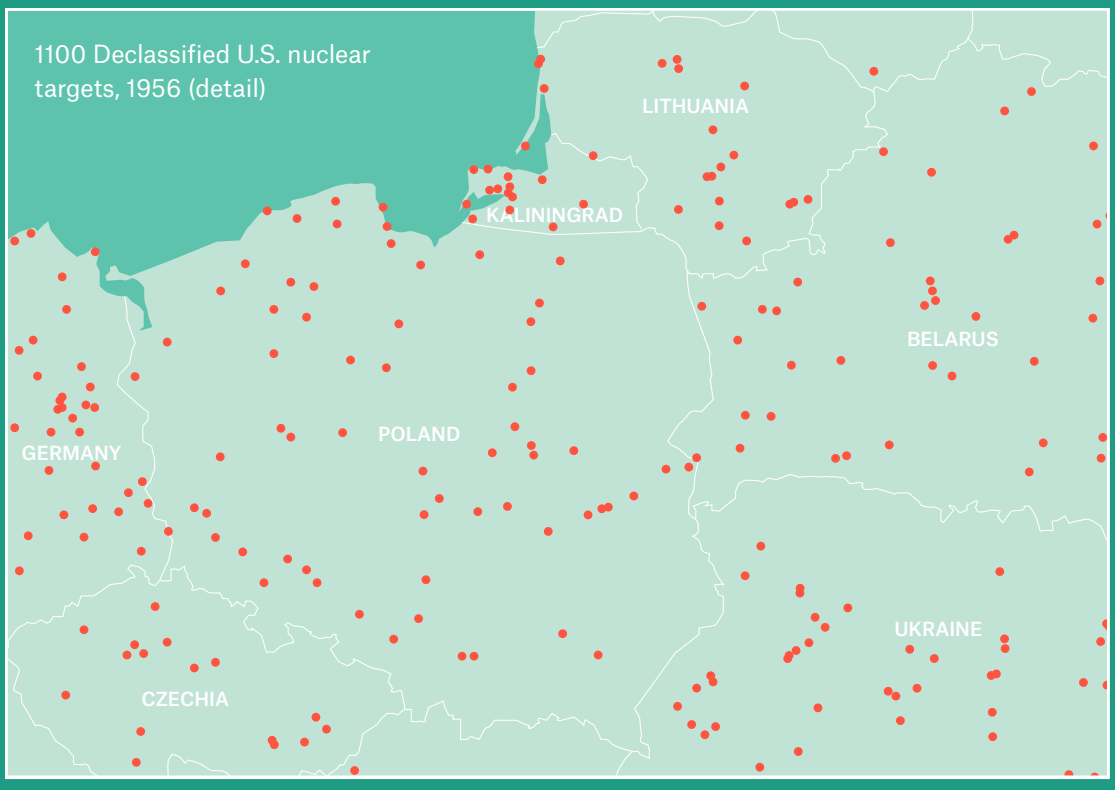
7. Unclassified summary of NAVWAG Study No. 5, Jan. 22, 1958, White House Office Files, Office of Staff Secretary, Subject Series, Alpha Subseries, Box 21, Nuclear Exchange (I) folder, Box 21, Eisenhower Library.

8. *Wizards of Armageddon*, 217-19, 237-45.

9. Ibid., Ch. 17.

1100 Declassified U.S. nuclear targets, 1956

From the National Security Archives



Source: https://futureoflife.org/background/us-nuclear-targets/?cn-reloaded=1&cn-reloaded=1&cn-reloaded=1

need to buy more than one new nuke. In other words, an unending arms race was a consequence of the policy — and the Soviets were now starting to deploy a fair number of nukes. So, in 1964, McNamara declared a new nuclear policy, stating — in very similar terms to the Navy's strategy of a few years earlier — that the U.S. needed only enough weapons to kill roughly 30 percent of the Soviet Union's population and half of its industrial capacity. At that point, all of the major cities would be destroyed; launching any more weapons would inflict only marginally more damage.

McNamara called this new policy "assured destruction." (A critic of the idea, who supported counterforce, lampooned it as "mutual assured destruction" in order to come up with the acronym MAD.)[10] But McNamara was being disingenuous. In a top secret memo to President Lyndon B. Johnson, McNamara wrote that MAD did not reflect how the U.S. would actually use nuclear weapons in a war. In fact, only a few hundred bombs and warheads out of the arsenal's several thousand were aimed at "urban-industrial" targets; the rest were aimed at military targets (though many were near or inside cities, so tens of millions of civilians would still die).

Counterforce was still the strategy; a plan for nuclear war fighting was always — secretly — an element of credible nuclear deterrence. MAD — the threat to destroy civilian targets in response to a Soviet first strike — was a political device to curb the Air Force's appetite.[11]

The generals, of course, were livid. McNamara left their strategy in place but, in their eyes, didn't give them enough weapons to execute it. They'd asked for 2,000 ICBMs, but McNamara capped the number at 1,000. So, the generals came up with their own clever scheme. NASA scientists were designing a rocket that, once in outer space, could unfurl several satellites into different orbits. Air Force scientists saw that they could adapt the technology so that an ICBM could fire several warheads at different targets. They called the program the MIRV (for Multiple Independently targetable Reentry Vehicle), and proposed installing MIRVs on an upgrade of the Minuteman missile, called the Minuteman III.

The generals told McNamara that they would accede to 1,000 Minuteman missiles if McNamara approved MIRV. McNamara took the deal.[12] The generals had outsmarted him. Over the next decade, they built 550 Minuteman IIIs, each with three warheads; along with their 500 single-warhead missiles, that gave them 2,150 ICBM warheads. In other words, the generals got what they wanted, plus some.

Soon, the Navy put MIRVs on its submarine-launched missiles — and, a bit later, they too built missiles accurate enough to destroy Soviet military targets, as a result of which they abandoned their "finite deterrence" philosophy.

10. Donald Brennan, "Strategic Alternatives: I," *New York Times*, May 24, 1971.

11. McNamara to Johnson, Recommended FY 1965-69 Strategic Retaliation Forces, Dec. 6, 1963, *Foreign Relations of the United States, 1961-63, Vol. VIII*, National Security Policy, Document 151.

12. McNamara to Johnson, Dec. 3, 1964, National Security Archives/Electronic Briefing Book #311, Document 2. (Originally in *Wizards*, 363-64.

Pulling Back the Curtain

All of this was highly classified. Most people, even those versed in nuclear strategy and history, thought that the U.S. policy was MAD. In 1974, during Gerald Ford's brief presidency, when Secretary of Defense James Schlesinger publicly announced a policy of "limited nuclear options," aiming nuclear weapons at Soviet *military* targets while avoiding cities, many politicians and analysts thought this was a new and dangerous departure from the philosophy of deterrence. But it was not at all new (nor was it understood that Schlesinger's most limited options would still involve firing a couple hundred nuclear weapons). The real policy was never MAD — it was always counterforce; Schlesinger's policy was a refinement of that policy. As long as an arms race was on, counterforce extended the race in perpetuity: More nukes meant more targets, more targets required more nukes.

But this is not to say that the impulse to build more was entirely rational.[13] In 1989, soon after George H.W. Bush was sworn in as president, his secretary of defense, Dick Cheney, was briefed on the latest version of the nuclear war plan. Cheney asked his assistant on strategic issues, a civilian analyst named Franklin Miller, to sit in. Miller had perused the array of classified documents reciting the rationales for limited nuclear options. Yet, he noticed, the briefing said nothing about such options.

Cheney and Miller were also struck by one detail in the war plan: It called for hitting the Soviet transportation network with 725 nuclear weapons. Cheney asked the briefer, a SAC general, why. The general shrugged and said he'd get back to him on that. (He never did.) After the meeting, Cheney told Miller to go out to Strategic Air Command's headquarters, in Omaha, and conduct a thorough review of the war plan; he alerted the officers at SAC that Miller should have full authority to look at everything.

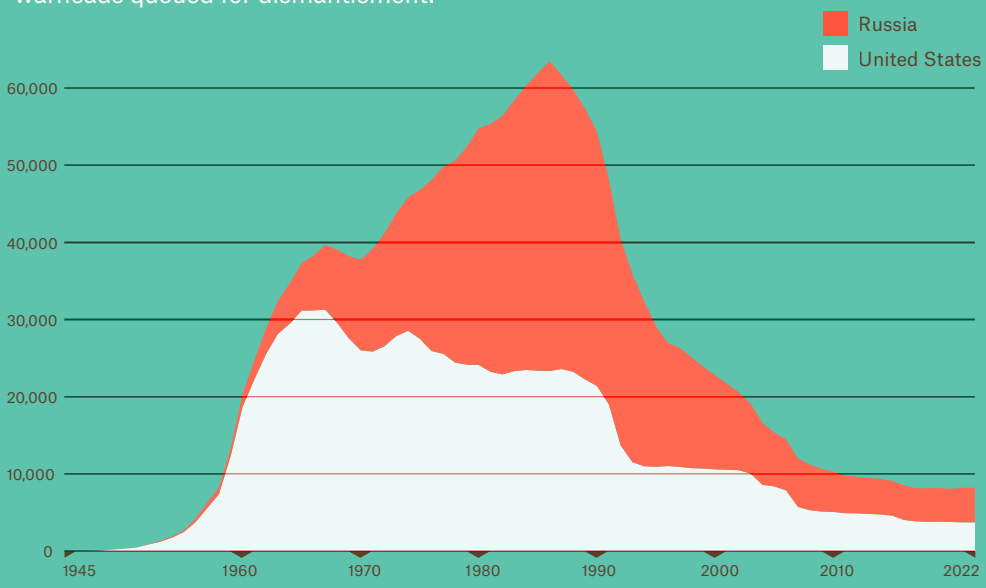
What Miller discovered made the term "overkill" seem a gross understatement. For example, just outside Moscow, the Soviets had an ancient anti-ballistic missile system holding 68 interceptors. After the Cold War, U.S. inspectors discovered that the system was completely useless. But the war plan specified that the site had to be destroyed with near-total certainty. SAC intelligence estimated (incorrectly) that each of the Soviet interceptors had a high probability of shooting down an incoming American warhead. So, JSTPS — Omaha's nuclear targeting agency — assigned 69 warheads to hit the site, to make absolutely certain that at least one of the warheads got through.

Another jaw-dropping example: One part of the nuclear war plan called for destroying the Soviet tank army. As a result, JSTPS aimed a lot of weapons at not only the tanks themselves, but also the factory that produced the tanks, the steel mill that supplied the factory, the ore-processing facility that supplied the steel mill, and the mine that furnished the ore. Miller and his staff learned that some SAC analysts had already pointed out the

13. The following is based almost entirely on interviews I conducted for *The Bomb*, Ch. 8.

Estimated nuclear warhead stockpiles, 1945 to 2022

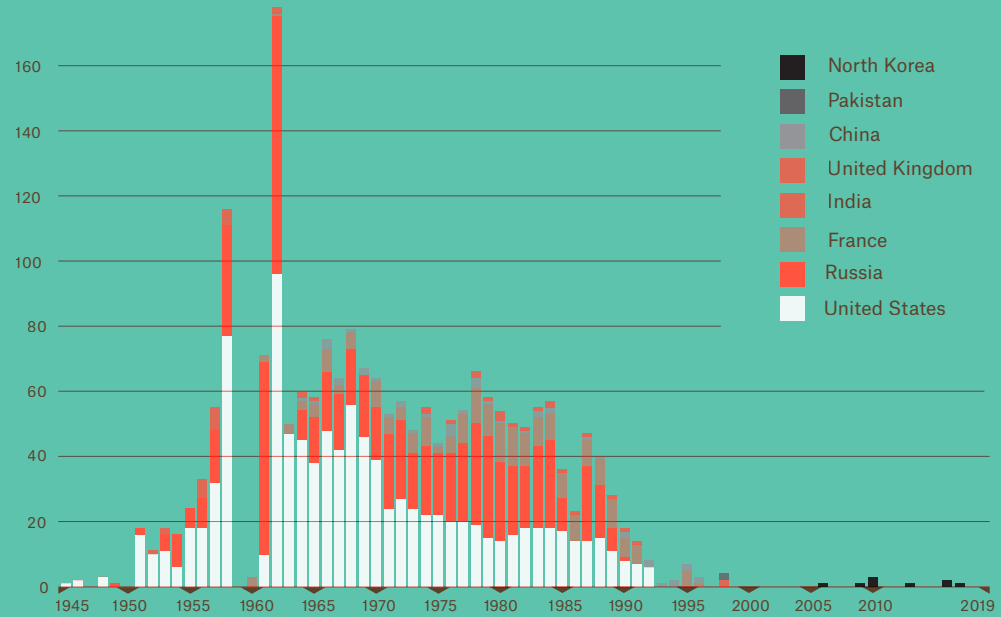
Stockpiles include warheads assigned to military forces, but exclude warheads queued for dismantlement.



Note: The exact number of countries' warheads is secret, and the estimates based on publicly available information, historical records, and occasional leaks. Warheads also vary substantially in their power.

Source: Federation of American Scientists (2022) / OurWorldInData.org/nuclear-weapons/

Number of nuclear weapons tests, 1945 to 2019



Source: Arms Control Association (2020) / OurWorldInData.org/nuclear-weapons/

excesses. A branch of math called nodal analysis suggested that, as long as the central links of a supply system were destroyed, there was no need to destroy every single piece; in many cases, just a few warheads, aimed at the right targets, would cripple the system. Gradually, Miller realized that the entire war plan was like this — a senseless aggregate of compartmentalized calculations.

Then came the key revelation. At this point the Bush administration was negotiating the Strategic Arms Reduction Treaty with the Soviets. During one of his trips, one of Miller's assistants asked a JSTPS officer whether the treaty's prospective cuts would affect SAC's ability to fulfill its mission — whether the U.S. could continue to deter nuclear war and limit damage if deterrence failed. The officer replied that he didn't do that sort of analysis. JSTPS, he went on, was prohibited from setting requirements or analyzing whether a certain kind of attack, with a certain number of weapons, would be militarily effective. When asked what the JSTPS actually did, the officer explained that they take all the weapons that are assigned to SAC and aim them at all the targets on the list.

The code was unlocked. It turned out that the war plan was based on supply, not demand — on how many weapons SAC *happened to have*, not on how many were needed.

Cheney ordered Miller and some officers in the Pentagon's Joint Chiefs of Staff — who were just as staggered by this revelation as the civilians — to go over every single weapon in SAC's arsenal and every single target on the JSTPS list with an eye toward figuring out how many nukes were really needed, even if the policy didn't change. This was the first time that any civilians — in fact, any official from Washington — truly scrutinized the nuclear war plan.

In the end, they calculated that, without any changes in the war plan's broad aims or policies, the arsenal could be cut in half — from roughly 12,000 to 5,888. After the Soviet Union imploded, they further reduced this figure, to around 2,200. (Much of this stemmed from removing targets in Eastern Europe, which now consisted of independent states, not Soviet allies.) As it happened, the Strategic Arms Reduction Talks (START) set 2,200 as the maximum number of weapons that each country could deploy. U.S. negotiators went into the talks knowing that the military chiefs — who ordinarily might protest such deep cuts — would be fine with them; and the chiefs were fine with such deep cuts because of the analysis forced by (of all people) Dick Cheney.

After the Cold War

For the first couple of decades of the post-Cold War era, few people thought, much less worried, about nuclear war. Presidents Bill Clinton and George W. Bush each had tangles with North Korea, but they didn't rise beyond the realm of rhetoric. Clinton and his team devoted time to pulling Boris Yeltsin's new Russia out of various crises, helping to secure the disheveled country's loose nukes and forging relations with the newly independent states of Eastern Europe, some of which were eager to join NATO. But tensions between the U.S. and Russia had abated; existing arms control treaties held firm. Remarkably, the size and structure of each side's nuclear arsenals — and their basic attitude toward nuclear weapons — held firm as well. President Barack

Obama did revisit the arsenal, yet even he held back from the bold steps that his analysis suggested were possible.

In 2010, Obama asked his top national-security staffers to conduct an analysis similar to Cheney's.[14] He had just signed the New START treaty with Russian President Dmitri Medvedev, which cut the arsenals further to 1,550. Eager to push ahead with still deeper cuts (one of his passionate aims, upon taking office, was to "reduce the role of nuclear weapons"), Obama wanted to know how deep the cuts could safely go.

Every couple of weeks for four months, a group of senior officials from the Pentagon and the National Security Council met with Gen. Robert Kehler, the head of Strategic Command (the new name for what had been SAC, referred to as StratCom), to go over, once again, every single weapon and every single target. They broke down the targets into categories and asked whether StratCom needed to hit every target or just some — and whether fewer weapons would be adequate to hit each target.

The leader of the Washington team, a Pentagon civilian named Jim Miller, was well aware of what Frank Miller (no relation) had done. He understood the practice, dating back to 1960, of pushing up the "required" number of weapons by elevating the certainty with which certain targets be destroyed. Kehler understood this too. If the war plan required that Target X be destroyed with 90 percent certainty, that meant the U.S. would need to launch two warheads against the target. If you're willing to reduce this probability to 75 percent, Kehler told the staffers, then one warhead would be sufficient. It was up to them, the political decision makers.

Even so, after all the paring down, Miller concluded — and Kehler concurred — that the U.S. could safely eliminate one-third of its nuclear arsenal; in other words, that New START's limits of 1,550 bombs and warheads could be further cut to 1,000. This too was a compromise. Miller reflected that even 1,000 left in place a lot of overkill; in the event of war, it meant merely that "the rubble would bounce one less time."

Then Kehler and the Joint Chiefs threw in a caveat: They would not publicly endorse a reduction of this magnitude unless the Russians cut their arsenal by the same amount in a follow-on treaty to New START.

The code was unlocked. It turned out that the war plan was based on supply, not demand — on how many weapons SAC happened to have, not on how many were needed.

14. The following is based almost entirely on interviews I conducted for *The Bomb*, Ch. 10.

Obama agreed. Though he understood that the nation would be perfectly secure with one-third fewer nuclear weapons, regardless of what Moscow did, he saw no point in making unilateral cuts. He wanted to negotiate deeper cuts with the Russians, who might see no need to go along if Washington made deep cuts on its own. He also knew he would get no support for unilateral cuts from Congress. Obama was a visionary, but when it came to policy, he was pragmatic. As one of his aides put it, he liked to "paint within the lines."

Not only did Obama make no further cuts, he got trapped into appearing to do the opposite. As part of a deal to get the Senate to ratify New START (which required a two-thirds majority), Obama signed a letter pledging that he would request funds to "modernize or replace" all three legs of the Triad (emphasis added).[15] This was a carefully written note, as "modernize" could mean upgrading a missile's software or installing new communications gear. Obama did not regard the letter as a promise to buy any new weapons; nor did he attach a dollar figure to the pledge.

However, the Republican hawks in Congress rolled out a list of all-new weapons and calculated that the package would cost $1.3 trillion over the next 30 years. (The estimate has since been revised to $2 trillion.) More than that, after the 2016 election, when Donald Trump came to office, newly appointed officials in the Pentagon, some of whom had worked for Republican senators, referred to this list as "the Obama program of record," thus ensuring that anyone who proposed a less ambitious plan would be tagged "more dovish than Obama."

Getting agreement on this required some sleight of hand. Trump's first secretary of defense, Jim Mattis, was a retired Marine general and therefore had little interest in nuclear weapons. (The Marines retired their short-range nukes decades ago and never built long-range weapons.) In fact, two years earlier, Mattis testified as a private citizen that it might be a good idea to get rid of the land-based ICBMs.[16] He had come under the influence of arms-control analysts, who argued that such weapons were inherently "destabilizing." They were at once the most accurate and the most vulnerable weapons; therefore, in a crisis, their very existence could provide an enemy with incentives or excuses to launch a preemptive first strike. Many years earlier, ICBMs had served a specific function: They were the only weapons that could quickly knock out blast-hardened targets, such as ICBM silos. However, since 1990, the Navy's Trident II submarine-launched ballistic missiles were accurate enough to destroy hardened targets as well. Land-based ICBMs had no purpose whatsoever.

But then someone (I have been unable to trace who) came up with a new rationale: *the sponge theory*.[17] The idea was that, if some president dismantled all the land-based ICBMs, there would be only five strategic targets in the continental U.S. — three bomber bases and two submarine ports. The Russians could launch an attack on those five targets with just one or two multiple independently

15. Message from the President on the New START Treaty, Feb. 2, 2011.

16. Mattis testimony, Senate Armed Services Committee, Jan. 27, 2015.

17. *The Bomb*, 274-75.

targetable reentry-vehicle missiles — and in a crisis, they might think they could pull it off. The U.S. would be left with only the half dozen or so submarines constantly at sea, and the president might not launch those missiles, knowing that if he or she did, the Russians could strike back with much more. On the other hand, the theorists went on, if the U.S. kept its 400 ICBMs, the Russians would have to fire 800 warheads to destroy them — and that would constitute a "major" attack. Any American president would have to retaliate, and the certainty of that prospect would deter the Russians from attacking in the first place.

This was a bizarre theory for three reasons. First, nuking just those five U.S. bases would kill hundreds of thousands, possibly millions, of American civilians; a president couldn't be counted on to do *nothing*. Second, during the Reagan years, some of these hawks argued that the Soviets would launch 2,000 warheads in order to destroy the 1,000 ICBMs that the U.S. had at the time, without blinking an eye, even though such an attack would kill tens of millions of Americans. Finally, allowing that the sponge theory might have some logic, it is extreme to contend that the U.S. needs 400 missiles to keep the Russians at bay; the sponge effect of 100 or even a dozen ICBMs would require the Russians to launch 200 or two dozen warheads respectively — certainly a "major" attack by any measure.

In any case, however many sponges we might need, it is doubtful that a whole new missile is necessary for the task. The present Minuteman missiles have been modified several times over the past 50 years, and they could be modified several times more.

The same is true of the B-52 and B-1 bombers, which can stay aloft and fire air-launched cruise missiles — which have a range of 1,500 miles, making it unnecessary for the airplanes to penetrate Russian airspace, where they might be shot down by air-defense missiles. A brand-new bomber isn't needed to launch cruise missiles.

New missile-launching submarines will be needed sometime in the future. Submarines are undetectable and, therefore, as long as nuclear weapons exist, they are a stabilizing force in the balance of terror. It is also a good idea to keep improving command-control-communications technology, to make sure that the weapons launch if the president wants to launch them (if just to maintain credible deterrence) and to minimize the chance of an unauthorized launch.

But in Congress, the White House and the civilian-run corridors of the Pentagon, the rationales for revamping all three legs of the Triad — and preserving the present numbers of bombers, missiles and warheads — gained support.

A Return to First Principles

There was a brief time, shortly before and just after the 2020 election, when some legislators, even some conservatives, began to think about putting some clamps on the nuclear juggernaut. COVID-19 and its economic side effects were soaking up hundreds of billions of tax dollars; a movement was afoot to rethink priorities, to wonder whether cuts were possible in other programs, including nuclear programs. But then came the upswing in Putin's aggressiveness in his decision to invade Ukraine, building on his interference with U.S. elections and his moves on Syria. Tensions

with China escalated as well. To argue against building more nuclear weapons might be seen as being "soft on defense."

At that point, *perceptions* overtook the slight steps toward objective analysis. This happened frequently during the Cold War. Whenever someone challenged the need to build new nuclear weapons, a general would reply that the U.S. cannot be "perceived" to be behind the enemy, even if objectively it wouldn't matter. The main argument these days for new nukes — and, to some, for a larger arsenal of nukes — is that they're needed because the Russians continue to upgrade their arsenal and the Chinese are expanding theirs. But just because the Russians and Chinese are wasting their money on more nuclear weapons than they need doesn't mean the Americans have to follow suit.

Adm. Charles Richard, the head of Strategic Command, said this past March that the U.S. needs to rethink the whole concept of deterrence because it now needs to deter two "near peer powers" — Russia and China — simultaneously.[18] This is questionable. First, the U.S. nuclear war plan has always taken into account the possibility of a war with more than one nuclear-armed country. Second, as Richard explicitly noted, China isn't quite a "peer power"; it has about one-tenth as many nuclear bombs and warheads as the United States or Russia.

China *is* a "peer power" in the sense that, if attacked with nuclear weapons, it could — even with its much smaller arsenal — retaliate with devastating consequences. But that only bolsters the argument that the U.S. could get by with far fewer weapons than it presently has. Back in the mid-1960s, the Kremlin considered attacking China, but Mao Zedong's paltry arsenal at the time — about a half dozen A-bombs — deterred Leonid Brezhnev from doing so. A few American presidents in recent decades have thought about attacking North Korea — but the Kim regime's handful of nukes has kept Washington at bay. The U.S. probably needs more than a few dozen nuclear weapons — but does it need a *lot* more than that? And if so, for what?

We would be better off mounting a full rebuttal of the vague assertions about "perceptions" (no one argues that some wasteful anti-poverty program should be funded anyway, so that it *looks* like we're fighting poverty). It's time for a return to first principles, the basic questions that nobody has been asking: How much is enough? And enough to do what? Enough material has been declassified over the last few decades that informed citizens can hold hearings and stage debates, even if members of Congress won't. At long last, let the hearings and debates begin.

18. Adm. Richards, testimony, House Armed Services Committee, Mar. 1, 2022.

Some of this article is based on research for my two books on nuclear issues, *The Wizards of Armageddon* (Simon & Schuster, 1983) and *The Bomb* (Simon & Schuster, 2020).

104

Dispatch from the Future

They May as Well Grow on Trees: The Future of Genetically Engineered Livestock

Xander Balwit

ILLUSTRATIONS BY
Natalya Balnova

On Aug. 18, 2053, Tyson Foods unveiled its much-anticipated product, Well Beef, at a benefit dinner in Lower Manhattan. Well Beef, a genetically engineered animal product derived from what the company is calling "Welfare-Enhanced Cows," is the third GE food product that Tyson has released and comes just a year on the heels of Ecopig.

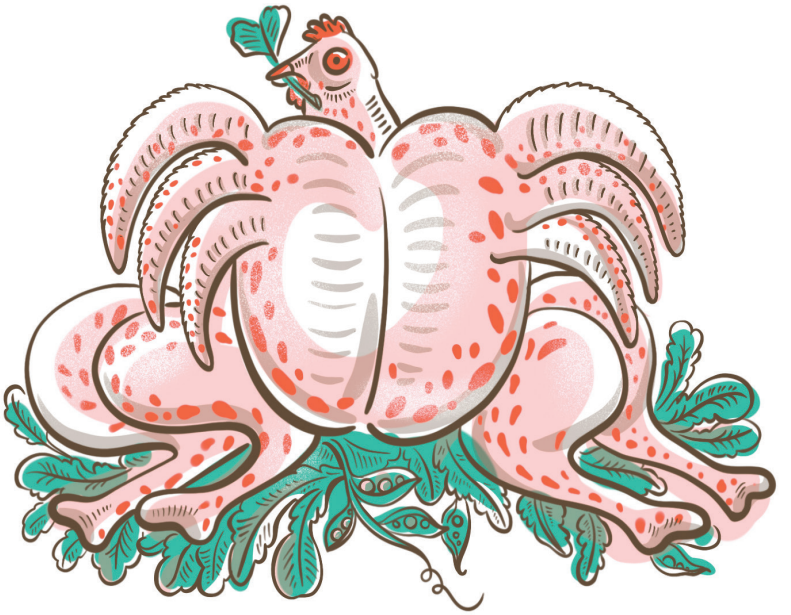
"It's exceptional!" exclaimed Grant Willis, the company's CEO, dabbing at his chin with a napkin. "We have finally achieved the Big Three. We have enjoyed phenomenal success with Pure Chicken and Ecopig, and now we are ushering in the era of Well Beef." Willis gestured toward the room, where guests eddied about the tables surveying and sampling Tyson's array of genetically engineered foods. The chefs had transformed the Well Beef into carpaccio woven into the shape of roses. Interspersed were silver trays of Ecopig sliders and Pure Chicken pâté nestled among garlands of fruit and salad greens.

As Pure Chicken celebrates half a decade of success this year, Well Beef is the first commercially available GE beef product that claims to be cruelty-free. "Genetic engineering for increased welfare" boasts a banner draped behind the podium. Light from the chandeliers shimmered across the marbled beef, giving the meat a dewy sheen. Lush bouquets of flowering sweet pea vines were placed throughout the rooms. "An homage to trait selection," Herbert Muller, an attending geneticist from Tyson, tells me. "On this day, it is worth reflecting on humanity's journey through the labyrinth of heredity."

Unlike these pea vines, whose differences are largely cosmetic, what makes Well Beef distinct from traditional meat goes beyond appearances. By using CRISPR and other gene editing technology, companies like Tyson can make targeted modifications to the genome in order to delete or insert new genes. While those who selected favorable variants in the past were doing so in ways that were molecularly indiscriminate, today's food engineers have gone from merely interpreting the gene to manipulating it.

This kind of genomic selection was first introduced into the dairy industry in 2009, where it was adopted rapidly. By 2015, more than half of all artificial insemination matings in the United States were made to genomically tested young bulls, resulting in cows that were larger and produced more milk. But even in this era, the gene was not being changed at its core, only selected for. It took decades of advancements in molecular technologies before this kind of targeted genomic selection could be made more efficient.

"When I was working in agricultural biotech back in the mid-2020s, things were different," says Vivian Carver, another attendee, a retired researcher at the Institute for AgriGenetic Welfare. "We



had the technology to edit the genome but regulatory gridlock had us in a chokehold. Every meal we have ever consumed is genetically distinct from every other meal and yet explicitly editing a gene made regulators pounce." For example, under the FDA's 2009 Guidance for Industry entitled "Regulation of Genetically Engineered Animals Containing Heritable rDNA Constructs," any organism that was modified with rDNA techniques was viewed as containing "drugs" and regulated as such; a single deletion in the genome triggered strict regulatory inspection.

This meant that GE research was subjected to strict oversight — even when the part of the animal being modified had no direct connection to the part of the animal consumed or producing food. For example, researchers attempting to shorten cattle horns to eliminate the need for painful dehorning practices were forced to present evidence that alterations in the horns would not affect the milk. Unsurprisingly, the milk was indistinguishable. Many scientists pushed back. After getting rejected for a grant back in 2019, former biotech researcher Alison van Eenennaam and her colleagues published their protestations in *Nature Magazine*, asking, "How can the absence of a small piece of DNA rationally be considered a drug?"

Other scientists, including Carver, simply left. Dismayed by the opportunity costs of adhering to regulatory restrictions, many joined or established research centers throughout Central and South America, where regulations on animal research were minimal or non-existent.

Thus, the first two decades of the 20th century were somewhat of a dark age for GE livestock, particularly in the United States. Only a few products like AquAdvantage Atlantic Salmon (which had already sat in regulatory limbo for over twenty years) were able to fight their way to commercial success. Many projects stalled and agriculture futures bottomed out. While GE animals floundered, yeast- and plant-based protein enjoyed their zenith. The popularity of plant-based diets exploded across wealthy nations, though the rate of overall meat consumption hardly dipped due to steadily growing wealth and populations in developing nations. Cultured meat also failed to take off in the way that advocates for alternative protein had hoped. Despite scientific advancements, scaling up cultured meat enterprises proved remarkably challenging and expensive.

However, in retrospect, the livestock revolution was beginning to happen quietly. Improvements geared toward the welfare of livestock and animal products began at the margins. Early advancements weren't conducted under the banner of animal welfare, but the most pressing public fears of the time: climate change and pandemics. Initial gene modifying efforts were directed at disease resistance or climatic adaptations. Early in the 2020s, agricultural researchers selectively bred cattle to have shorter hair to help them regulate heat and engineered PRRS-virus-resistant pigs. Emboldened by these successes, scientists continued to push the capabilities of genetic engineering even as they faced funding constraints and regulatory overreach.

It was in the 2030s that the tide truly began to change. Efforts to abolish intensive animal agriculture continued to fail. States that tried to shift their agricultural production away from livestock faced

exacting retribution. After the court ruled in favor of concentrated animal feeding operations (CAFOs) in the landmark 2037 court decision Cal-Maine Foods, Inc. v. Iowa, it appeared that the meat lobby was too strong. However, their influence was soon to be curtailed. After a wave of zoonotic pandemics such as H1N1-39 broke out in the 2040s, the National Pork Producers Council and their subsidiaries embarked on a frenzied crusade to rehabilitate their public image, ultimately shaping the trajectory of today's livestock industry.

Facing unrelenting political pressure stoked by the public's concern over zoonotic pathogens, CAFO lobbies were forced to compromise. In a series of unprecedented negotiations, companies like Cal-Maine and Tyson sat down with advocates from the Humane League and the National Institute of Allergy and Infectious Disease to discuss the future of meat production. If consumers and corporations refused to give up meat entirely, there had to be another solution for those demanding animal welfare and human safety.

The upshot was monumental: CAFOs could continue to operate as long as they worked toward disease and pathogen resistance — a determination that called upon genetic engineering as an engine of compromise.

"Although for many it was sad to see technology succeed where moral arguments had not, it seemed the second-best solution," conceded bioethicist and historian Caldwell Marin during an interview. "The meat industry was placated, and those who cared about the welfare of humans and animals were emboldened to make real change."

For decades, the rift between those working within animal agriculture and those fighting for an overhaul of factory farming led both sides to weaponize gene-editing technology against each other. The narrative was that GE could either be aimed at efficiency *or* at welfare, but never at both. However, the incompatibility of efficiency and ethics proved false, and once these interests were aligned, the necessary technology emerged rapidly.

The 2040s brought about the GE renaissance. Fearful of the opportunity costs of delaying transgenically modified organisms after the post-pandemic negotiations of the '40s, CAFOs increased their pressure on regulatory bodies like the FDA. This pressure, alongside the bold promises for pathogenetically resistant livestock by the National Institute of Allergy and Infectious Disease, finally prompted the FDA to initiate a regulatory overhaul. With it came a deluge of talent and resources, and it wasn't long after that researchers like Carver flocked back to the United States. Still held to rigorous standards with respect to toxicity and allergenicity, the language and simplified reporting specifications in the new Guidance for Industry

allowed for easier R&D. Finally unimpeded, biotech researchers could push the boundaries of genetic engineering — from changing an animal's experience of heat to altering its ability to experience entirely.

Tyson's Pure Chicken was the first GE animal engineered not to perceive pain. Using CRISPR and other proprietary technologies, bioengineers were able to manipulate the chickens so they had brain function sufficient for maintaining growth but not for supporting mental states or psychological experiences. These chickens, which lacked beaks, eyes and feathers, also had ablations to their anterior

cingulate that disrupted the affective dimensions of pain. Their secondary somatosensory cortex was left intact, rendering them able to eat and drink, and even to react instinctually to stimuli. But when exposed to adverse stimuli, rather than exhibiting nociceptive behavior, they remained serene. They resembled something between an animal and a fruit, an observation that is encapsulated by the product's official slogan: "They may as well grow on trees."

Although the product had its critics, its immediate commercial success left no doubt about the industry's trajectory. Within months of its debut in 2048, Pure Chicken became the industry standard. Non-genetically engineered chicken simply could not compete with

Pure Chicken for taste or efficiency. And whereas traditionally bred chickens are prone to pecking one another's eyes out when too tightly confined, Pure Chickens are equanimous and placid. From temperament to taste, cruelty-free chicken outmatched non-GE poultry.

In the ebullience that followed the commercial success of Pure Chicken, companies like Tyson and Cal-Maine Foods turned their attention to bioengineering a larger array of more complex livestock animals. The methods that researchers used were similar, focusing on disrupting the neural pathways so that they could alleviate pain while not stunting growth. Three years ago, in the spring of 2050, a team of animal science researchers from California Polytechnic State University discovered that folic acid deficiency during embryogenesis could lead to a neural tube defect that disrupts pain signaling in the brain. They first implemented this strategy in pigs, which led to a spate of GE pork products, including Ecopig, before turning their attention to modifying cattle.

Well Beef is thus the tour de force of GE livestock. The welfare-enhanced cows from which Well Beef is manufactured are a genetic hybrid of Holstein and Angus cattle. Large and muscle-bound, their architectonic bodies ripple with prime cuts. The most pronounced distinction between these beef cows and their forebears is their heads, which develop with a concave brain but retain a partial skull, including the face.

According to Tyson, these cattle eat, grow, live and die without a vestige of pain. Even the most skeptical evaluators confirmed this appraisal. Upon visiting Tyson's headquarters last month, Maxwell Harder, an investigator with the Factory Farming Awareness Coalition, marveled that he had "plausibly borne witness to the largest reduction of suffering ever undertaken."

But some neuroscientists and bioethicists remain skeptical.

"The brain is an astonishingly complex organ," said neuroscientist Masha Ruhig of the Center for Neuroscience and Society. "Millions of neurons throughout our brain signal pain. These neurons are spread about in every region. Many of my colleagues are concerned that once we ablate the regions of the brain that are connected to the conscious recognition of pain that the brain will simply reforge these connections elsewhere."

In most cases, scientists celebrate the brain's ability to compensate for damage. But in the case of GE livestock, this kind of neuroplasticity could put an end to the industry's optimism. The billions of farm animals undergoing genetic modifications so that they cannot experience suffering might still feel pain in a manner that evades our current understanding of neuroscience. Researchers and

ethicists like Ruhig fear that with the absence of a fully formed brain, the neural signals of these GE animals will simply reemerge in different regions. In doing so, neural pain responses might evade our detection.

In other words, genetically engineered animals might still experience pain because bioengineers are either wrong about how to disrupt the pain response or because the conscious recognition of pain takes place somewhere unanticipated.

"I certainly don't want to be a prophet of doom," continued Ruhig. "If the bioengineers at places like Tyson are correct and these pain receptors really don't migrate, then what they have achieved is epoch shaping." If they're wrong, the industry will find itself not far from where it started in terms of the moral implications.

I recall her solemnity as the benefit dinner concludes with rousing addresses from Willis and others from Tyson outlining the turbulent history of genetically engineered livestock and speaking with rapture about developments to come. Alan Park, the company's program director, passionately details a pilot program geared toward growing animal limbs from a central node containing a slurry of nutrients and DNA. "Think of it as a vertical bestiary!" he says, smiling.

"Animal agriculture need not be stuck in the past," Park continues. "And thanks to our scientists' groundbreaking achievements, the contradiction between eating meat and eating ethically has vanished." His address evokes deafening applause.

A gentle breeze blows through the dining room, rustling the chandeliers and sending napkins up into the air. Park steps down from the lectern and osmoses into the audience between a flurry of handshakes. A waiter emerges beside me and offers me a tray of Well Beef. I take a sample, dropping the parsley garnish into my napkin.

The meat melts against the roof of my mouth, marbled, succulent. It might be my imagination, but I can taste the sun-baked pastures of middle America, lands no longer grazed upon by sufferers, but by the beneficiaries of scientific advancement.

Scott Alexander is a writer and psychiatrist based in Oakland, California. He blogs at astralcodexten.substack.com.

Xander Balwit is a writer and wildland firefighter from Portland, Oregon. Unremittingly interested in what kind of future will befall us, Xander explores the plausible, dismal, and hopeful in her work.

Christopher Leslie Brown teaches history at Columbia University. He is the author of *Moral Capital: Foundations of British Abolitionism*, which was awarded the 2007 Frederick Douglass Book Prize.

Karson Elmgren is a Research Analyst at Georgetown's Center for Security and Emerging Technology (CSET), where he works on the AI Assessment team.

Kevin Esvelt is an assistant professor at the MIT Media Lab, where he leads the Sculpting Evolution Group in studying how ecosystems evolve and advancing biotechnology safely.

Fred Kaplan is the "War Stories" columnist for Slate and the author of six books, including *The Wizards of Armageddon*, *The Insurgents* (a Pulitzer Prize finalist) and *The Bomb*.

Jared Leibowich is a forecaster for Samotsvety and the Swift Centre. He placed 1st out of 7,000 for the *In the News 2021* Good Judgment competition and is currently ranked 1st for the *In the News 2022* competition. He can be reached at jleibowich@gmail.com.

Kelsey Piper is a senior writer at Vox's Future Perfect. She writes about emerging technologies, global development, pandemics, effective altruism, and what it'll take to make it safely to the 22nd century.

Stuart Ritchie is a senior lecturer in psychiatry at King's College London. He is the author of a book, *Science Fictions: Exposing Fraud, Bias, Negligence and Hype in Science*, and blogs about metascience at stuartritchie.substack.com.

Dietrich Vollrath is a professor and chair of the department of economics at the University of Houston. His work focuses on economic growth. He blogs at growthecon.com.