



Forecasting Newsletter

1. [Forecasting Newsletter: April 2020](#)
2. [Forecasting Newsletter: May 2020.](#)
3. [Forecasting Newsletter: June 2020.](#)
4. [Forecasting Newsletter: July 2020.](#)
5. [Forecasting Newsletter: August 2020.](#)
6. [Forecasting Newsletter: September 2020.](#)
7. [Forecasting Newsletter: October 2020.](#)
8. [Forecasting Newsletter: November 2020](#)
9. [Forecasting Newsletter: December 2020](#)
10. [2020: Forecasting in Review.](#)
11. [Forecasting Newsletter: January 2021](#)
12. [Forecasting Newsletter: February 2021](#)
13. [Forecasting Newsletter: March 2021](#)
14. [Forecasting Newsletter: April 2021](#)
15. [Forecasting Newsletter: May 2021](#)
16. [Forecasting Newsletter: June 2021](#)
17. [Forecasting Newsletter: July 2021](#)
18. [Forecasting Newsletter: August 2021](#)
19. [Forecasting Newsletter: September 2021.](#)
20. [Forecasting Newsletter: October 2021.](#)
21. [Forecasting Newsletter: November 2021](#)
22. [Forecasting Newsletter: December 2021](#)
23. [Forecasting Newsletter: Looking back at 2021](#)
24. [Forecasting Newsletter: January 2022](#)
25. [Forecasting Newsletter: February 2022](#)
26. [Forecasting Newsletter: April 2022](#)
27. [Forecasting Newsletter: March 2022](#)

Forecasting Newsletter: April 2020

A forecasting digest with a focus on experimental forecasting.

- You can sign up [here](#).
- You can also see this post on the Effective Altruism Forum [here](#)
- And the post is archived [here](#)

The newsletter itself is experimental, but there will be at least five more iterations. Feel free to use this post as a forecasting open thread.

Conflict of interest: With Foretold in general and Jacob Laguerros in particular. This is marked as (c.o.i) throughout the text.

Index

- Prediction Markets & Forecasting platforms.
 - Augur.
 - PredictIt & Election Betting Odds.
 - Replication Markets.
 - Coronavirus Information Markets.
 - Foretold. (c.o.i).
 - Metaculus.
 - Good Judgement Inc. & Good Judgement Open.
- In the News.
- Long Content.

Prediction Markets & Forecasting platforms.

Forecasters may now choose to forecast any of the four horsemen of the Apocalypse: Death, Famine, Pestilence and War.

Augur: augur.net

Augur is a decentralized prediction market. It will be undergoing its [first major update](#).

Predict It & Election Betting Odds: [predictIt.org](http://predictit.org) & electionBettingOdds.com

PredictIt is a prediction platform restricted to US citizens or those who bother using a VPN. [Anecdotally](#), it often has free energy, that is, places where one can earn money by having better probabilities, and where this is not too hard. However, due to fees & the hassle of setting it up, these inefficiencies don't get corrected. In PredictIt, the [world politics](#) section...

- gives a 17% to [a Scottish independence referendum](#) (though read the fine print).
- gives 20% to [Netanyahu leaving before the end of the year](#)
- gives 64% to [Maduro remaining President of Venezuela before the end of the year](#).

The question on [which Asian/Pacific leaders will leave office next?](#) also looks like it has a lot of free energy, as it overestimates low probability events.

[Election Betting Odds](#) aggregates PredictIt with other such services for the US presidential elections.

Replication Markets: [replicationmarkets.com](#)

Replication Markets is a project where volunteer forecasters try to predict whether a given study's results will be replicated with high power. Rewards are monetary, but only given out to the top N forecasters, and markets suffer from sometimes being dull. They have added [two market-maker bots](#) and commenced and conclude their 6th round. They also added a sleek new widget to visualize the price of shares better.

Coronavirus Information Markets: [coronainformationmarkets.com](#)

For those who want to put their money where their mouth is, there is now a prediction market for coronavirus related information. The number of questions is small, and the current trading volume started at \$8000, but may increase. Another similar platform is [waves.exchange/prediction](#), which seems to be just a wallet to which a prediction market has been grafted on.

Unfortunately, I couldn't make a transaction in these markets with ~30 mins; the time needed to be included in an ethereum block is longer and I may have been too stingy with my gas fee.

Foretold: [foretold.io \(c.o.i\)](#)

Foretold is an forecasting platform which has experimentation and exploration of forecasting methods in mind. They bring us:

- A new [distribution builder](#) to visualize and create probability distributions.
- Forecasting infrastructure for [epidemicforecasting.org](#).

Metaculus: [metaculus.com](#)

Metaculus is a forecasting platform with an active community and lots of interesting questions. They bring us a series of tournaments and question series:

- [The Ragnarök question series on terrible events](#)
- [Pandemic and lockdown series](#)
- [The Lightning Round Tournament: Comparing Metaculus Forecasters to Infectious Disease Experts.](#) "Each week you will have exactly 30 hours to lock in your prediction on a short series of important questions, which will simultaneously be posed to different groups of forecasters. This provides a unique opportunity to directly compare the Metaculus community prediction with other forecasting methods." Furthermore, Metaculus swag will be given out to the top forecasters.
- [Overview of Coronavirus Disease 2019 \(COVID-19\) forecasts](#).
- [The Salk Tournament for coronavirus \(SARS-CoV-2\) Vaccine R&D.](#)

- [Lockdown series: when will life return to normal-ish?](#)

Good Judgement Inc. & Good Judgement Open.

Good Judgement Inc. is the organization which grew out of Tetlock's research on forecasting, and out of the Good Judgement Project, which won the [IARPA ACE forecasting competition](#), and resulted in the research covered in the *Superforecasting* book.

The Open Philanthropy Project has funded [this covid dashboard](#) by their (Good Judgement Inc.'s) Superforecasting Analytics Service, with predictions solely from superforecasters; see more on [this blogpost](#).

Good Judgement Inc. also organizes the Good Judgement Open ([gjopen.com](#)) [<https://www.gjopen.com/>], a forecasting platform open to all, with a focus on serious geopolitical questions. They structure their questions in challenges, to which they have recently added one on [the Coronavirus Outbreak](#); some of these questions are similar in spirit to the short-fuse Metaculus Tournament.

Of the questions which have been added recently to the Good Judgment Open, the crowd [doesn't buy](#) that Tesla will release an autopilot feature to navigate traffic lights, despite announcements to the contrary. Further, the aggregate...

- is extremely confident that, [before 1 January 2021](#), the Russian constitution will be amended to allow Vladimir Putin to remain president after his current term.
- gives a lagging estimate of 50% on [Benjamin Netanyahu ceasing to be the prime minister of Israel before 1 January 2021](#).
- and 10% for [Nicolás Maduro](#) leaving before the 1st of June.
- [forecasts famine](#) (70%).
- Of particular interest is that GJOpen didn't see the upsurge in tests (and thus positives) in the US until until the day before they happened, for [this question](#). Forecasters, including superforecasters, went with a linear extrapolation from the previous n (usually 7) days. However, even though the number of cases looks locally linear, it's also globally exponential, as [this 3Blue1Brown video](#) shows. On the other hand, an enterprising forecaster tried to fit a Gompertz distribution, but then fared pretty badly.

In the News

- [Forecasts in the time of coronavirus](#): The Financial times runs into difficulties trying to estimate whether some companies are overvalued, because the stock value/earnings ratio, which is otherwise an useful tool, is going to infinity as earnings go to 0 during the pandemic.
- [Predictions are hard, especially about the coronavirus](#): Vox has a short and sweet article on the difficulties of prediction forecasting; of note is that epidemiology experts are not great predictors.
- [538: Why Forecasting COVID-19 Is Harder Than Forecasting Elections](#)
- [COVID-19: Forecasting with Slow and Fast Data](#). A short and crisp overview by the Federal Reserve Bank of St Louis on lagging economic measurement instruments, which have historically been quite accurate, and on the faster instruments which are available right now. Highlight: "As of March 31, the WEI [a faster, weekly economic index] indicated that GDP would decline by 3.04% at an

annualized rate in the first quarter, a much more sensible forecast than that which is currently indicated by the ENI (a lagging measure which predicts 2.26% growth on an annualized basis in the first quarter)".

- [Decline in aircraft flights clips weather forecasters' wings](#): Coronavirus has led to reduction in number of aircraft sending data used in making forecasts.
- [The World in 2020, as forecast by The Economist](#). The Brookings institution looks back at forecasts for 2020 by *The Economist*.
- Forbes brings us this [terrible, terrible opinion piece](#) which mentions Tetlock, goes on about how humans are terrible forecasters, and then predicts that there will be no social changes because of covid with extreme confidence.
- [The Challenges of Forecasting the Spread and Mortality of COVID-19](#). The Heritage foundation brings us a report with takeaways of particular interest to policymakers. It has great illustrations of how the overall mortality changes with different assumptions. Note that criticisms of and suggestions for the current US administration are worded kindly, as the Heritage Foundation is a conservative organization.
- [Why most COVID-19 forecasts were wrong](#). Financial review article suffers from hindsight bias.
- [Banks are forecasting on gut instinct — just like the rest of us](#). Financial Times article starts with "We all cling to the belief that somebody out there, somewhere, knows what the heck is going on. Someone — well-connected insider, evil mastermind — must hold the details on the coming market crash, the election in November, or when the messiah will return. In moments of crisis, this delusion tightens its grip," and it only gets better.
- ['A fool's game': 4 economists break down what it's like forecasting the worst downturn since the Great Recession](#). "'My outlook right now is that I don't even have an outlook,' Martha Gimbel, an economist at Schmidt Futures, told Business Insider. 'This is so bad and so unprecedented that any attempt to forecast what's going to happen here is just a fool's game.'"
- [IMF predicts -3% global depression](#). "Worst Economic Downturn Since the Great Depression".
- [COVID-19 Projections](#): A really sleek US government coronavirus model. See [here](#) for criticism. See also: [Epidemic Forecasting](#) (c.o.i).
- [The M5 competition is ongoing](#).
- [Some MMA forecasting](#). The analysis surprised me; it could well have been a comment in a GJOpen challenge.
- [Self-reported COVID-19 Symptoms Show Promise for Disease Forecasts](#). "Thus far, CMU is receiving about one million responses per week from Facebook users. Last week, almost 600,000 users of the Google Opinion Rewards and AdMob apps were answering another CMU survey each day."
- [Lockdown Policy and Disease Eradication](#). Researchers in India hypothesize on what the optimal lockdown policy may be.
- [Using a delay-adjusted case fatality ratio to estimate under-reporting](#).
- [The first modern pandemic](#). In which Bill Gates names covid-SARS "Pandemic I" and offers an informed overview of what is yet to come.
- [36,000 Missing Deaths: Tracking the True Toll of the Coronavirus Crisis](#).
- There is a shadow industry which makes what look to be really detailed reports on topics of niche interest: Here is, for example, a [\\$3,500 report on market trends for the Bonsai](#)
- [An active hurricane season will strain emergency response amid pandemic, forecasters warn](#). "Schlegelmilch stresses that humanity must get better at prioritizing long-term strategic planning."

Long Content

- [Atari, early.](#) "Deepmind announced that their Agent57 beats the 'human baseline' at all 57 Atari games usually used as a benchmark."
- [A failure, but not of prediction](#); a SlateStarCodex Essay.
- [Philip E. Tetlock on Forecasting and Foraging as a Fox](#); an interview with Tyler Cowen. Some highly valuable excerpts on counterfactual reasoning. Mentions [this program](#) and [this study](#), on the forefront of knowledge.
- [Assessing Kurzweil's 1999 predictions for 2019](#). Kurzweil made on the order of 100 predictions for 2019 in his 1999 book *The Age of Spiritual Machines*. How did they fare? We'll find out, next month.
- [Zvi on Evaluating Predictions in Hindsight](#). A fun read. Of course, the dissing of Scott Alexander's prediction is fun to read, but I really want to see how a list of Zvi's predictions fares.
- An oldie related to the upcoming US elections: [Which Economic Indicators Best Predict Presidential Elections?](#), from 2011's Nate Silver.
- [A rad comment exchange at GJOpen](#) in which cool superforecaster @Anneinak shares some pointers.
- [As the efficient markets hypothesis turns 50, it is time to bin it](#) for a Financial Times article, from Jan 1st and thus untainted by coronavirus discussion.
Related: [This LW comment by Wei Dai](#) and [this tweet](#) from Eliezer Yudkowsky.
See also a very ramby article by an Australian newspaper: [Pandemic highlights problems with efficient-market hypothesis](#).

Forecasting Newsletter: May 2020.

A forecasting digest with a focus on experimental forecasting. The newsletter itself is experimental, but there will be at least four more iterations. Feel free to use this post as a forecasting open thread; feedback is welcome.

- You can sign up [here](#).
- You can also see this post on the EA Forum [here](#)
- And the post is archived [here](#)

Index

- Prediction Markets & Forecasting platforms.
 - Augur.
 - Coronavirus Information Markets.
 - CSET: Foretell.
 - Epidemic forecasting (c.o.i).
 - Foretold. (c.o.i).
 - /(Good Judgement?[^\n]*)(Superforecast(ing|er))/gi
 - Metaculus.
 - PredictIt & Election Betting Odds.
 - Replication Markets.
- In the News.
- Grab bag.
- Negative examples.
- Long Content.

Prediction Markets & Forecasting platforms.

Augur: augur.net

Augur is a decentralized prediction market. [Here](#) is a fine piece of reporting outlining how it operates and the road ahead.

Coronavirus Information Markets: coronainformationmarkets.com

For those who want to put their money where their mouth is, this is a prediction market for coronavirus related information.

Making forecasts is tricky, so would-be-bettors might be better off pooling their forecasts together with a technical friend. As of the end of this month, the total trading volume of active markets sits at \$26k+ (upwards from \$8k last month), and some questions have been resolved already.

Further, according to their FAQ, participation from the US is illegal: "*Due to the US position on information markets, US citizens and residents, wherever located, and anyone physically present in the USA may not participate in accordance with our Terms.*" Nonetheless, one might take the position that the US legal framework on information markets is so dumb as to be illegitimate.

CSET: Foretell

The [Center for Security and Emerging Technology](#) is looking for (unpaid, volunteer) forecasters to predict the future to better inform policy decisions. The idea would be that as emerging technologies pose diverse challenges, forecasters and forecasting methodologies with a good track record might be a valuable source of insight and advice to policymakers.

One can sign-up on [their webpage](#). CSET was previously funded by the [Open Philanthropy Project](#); the grant writeup contains some more information.

Epidemic Forecasting: [epidemicforecasting.org](#) (c.o.i)

As part of their efforts, the Epidemic Forecasting group had a judgemental forecasting team that worked on a variety of projects; it was made up of forecasters who have done well on various platforms, including a few who were official Superforecasters.

They provided analysis and forecasts to countries and regions that needed it, and advised a vaccine company on where to locate trials with as many as 100,000 participants. I worked a fair bit on this; hopefully more will be written publicly later on about these processes.

They've also been working on a mitigation calculator, and on a dataset of COVID-19 containment and mitigation measures.

Now they're looking for a project manager to take over: see [here](#) for the pitch and for some more information.

Foretold: [foretold.io](#) (c.o.i)

I personally added a distribution drawer to the [Highly Speculative Estimates](#) utility, for use within the Epidemic Forecasting forecasting efforts; the tool can be used to draw distributions and send them off to be used in Foretold. Much of the code for this was taken from Evan Ward's open-sourced [probability.dev](#) tool.

/(Good Judgement?[^*])|(Superforecast(ing|er))/gi

(The title of this section is a [regular expression](#), so as to accept only one meaning, be maximally unambiguous, yet deal with the complicated corporate structure of Good Judgement.)

Good Judgement Inc. is the organization which grew out of Tetlock's research on forecasting, and out of the Good Judgement Project, which won the [JARPA ACE forecasting competition](#), and resulted in the research covered in the *Superforecasting* book.

Good Judgement Inc. also organizes the Good Judgement Open [gjopen.com](#), a forecasting platform open to all, with a focus on serious geopolitical questions. They structure their questions in challenges. Of the currently active questions, here is a selection of those I found interesting (probabilities below):

- [Before 1 January 2021, will the People's Liberation Army \(PLA\) and/or People's Armed Police \(PAP\) be mobilized in Hong Kong?](#)
- [Will the winner of the popular vote in the 2020 United States presidential election also win the electoral college?](#)- This one is interesting, because it has infrequently gone the other way historically, but 2/5 of the last USA elections were split.
- [Will Benjamin Netanyahu cease to be the prime minister of Israel before 1 January 2021?](#) Just when I thought he was out, he pulls himself back in.
- [Before 28 July 2020, will Saudi Arabia announce the cancellation or suspension of the Hajj pilgrimage, scheduled for 28 July 2020 to 2 August 2020?](#)
- [Will formal negotiations between Russia and the United States on an extension, modification, or replacement for the New START treaty begin before 1 October 2020?](#)s

Probabilities: 25%, 75%, 40%, 62%, 20%

On the Good Judgement Inc. side, [here](#) is a dashboard presenting forecasts related to covid. The ones I found most worthy are:

- [When will the FDA approve a drug or biological product for the treatment of COVID-19?](#)
- [Will the US economy bounce back by Q2 2021?](#)
- [What will be the U.S. civilian unemployment rate \(U3\) for June 2021?](#)
- [When will enough doses of FDA-approved COVID-19 vaccine\(s\) to inoculate 25 million people be distributed in the United States?](#)

Otherwise, for a recent interview with Tetlock, see [this podcast](#), by Tyler Cowen.

Metaculus: [metaculus.com](#)

Metaculus is a forecasting platform with an active community and lots of interesting questions. In their May pandemic newsletter, they emphasized having "all the benefits of a betting market but without the actual betting", which I found pretty funny.

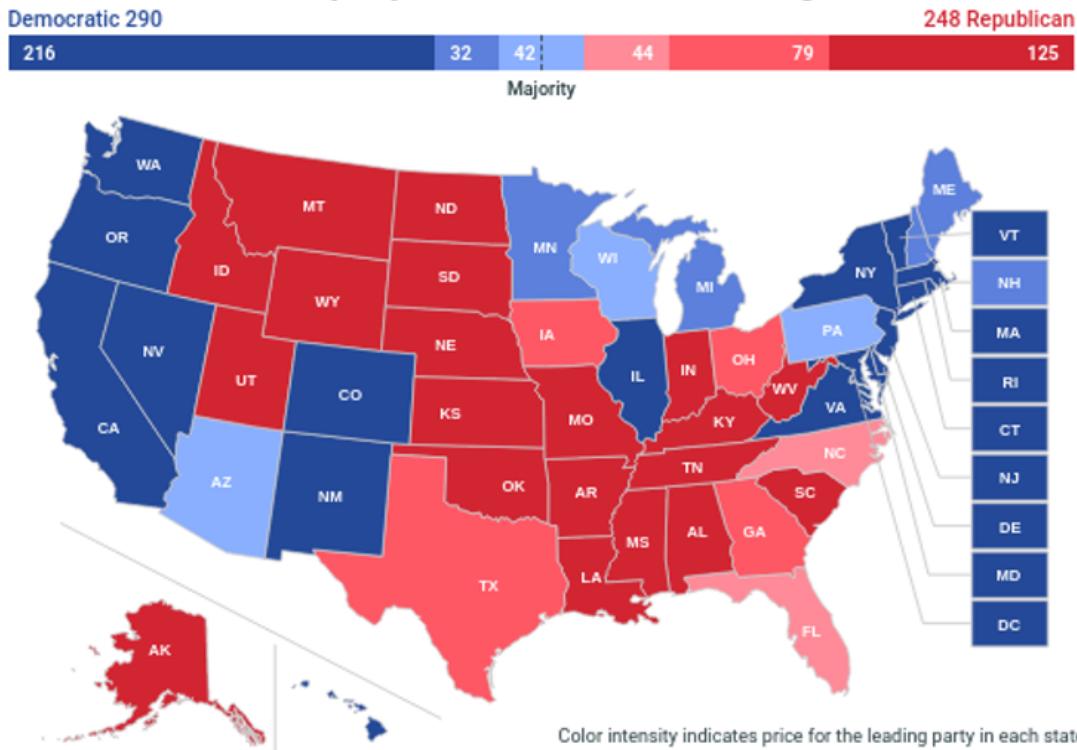
This month they've organized a flurry of activities, most notably:

- [The Salk Tournament](#) on vaccine development
- [The El Paso Series](#) on collaboratively predicting peaks.
- [The Lightning Round Tournament](#), in which metaculus forecasters go head to head against expert epidemiologists.
- They also present a [Covid dashboard](#).

Predict It & Election Betting Odds: [predictIt.org](#) & [electionBettingOdds.com](#)

PredictIt is a prediction platform restricted to US citizens, but also accessible with a VPN. This month, they present a map about the electoral college result in the USA. States are colored according to the market prices:

Which party will win the Electoral College?



Some of the predictions I found most interesting follow. The market probabilities can be found below; the engaged reader might want to write down their own probabilities and then compare.

- [Will Benjamin Netanyahu be prime minister of Israel on Dec. 31, 2020?](#)
- [Will Trump meet with Kim Jong-Un in 2020?](#)
- [Will Nicolás Maduro be president of Venezuela on Dec. 31, 2020?](#)
- [Will Kim Jong-Un be Supreme Leader of North Korea on Dec. 31?](#)
- [Will a federal charge against Barack Obama be confirmed before November 3?](#)

Some of the most questionable markets are:

- [Will Trump switch parties by Election Day 2020?](#)
- [Will Michelle Obama run for president in 2020?](#)
- [Will Hillary Clinton run for president in 2020?](#)

Market probabilities are: 76%, 9%, 75%, 82%, 8%, 2%, 6%, 11%.

[Election Betting Odds](#) aggregates PredictIt with other such services for the US presidential elections, and also shows an election map. The creators of the webpage used its visibility to promote [ftx.com](#), another platform in the area, whose webpage links to effective altruism and mentions:

FTX was founded with the goal of donating to the world's most effective charities. FTX, its affiliates, and its employees have donated over \$10m to help save lives, prevent suffering, and ensure a brighter future.

Replication Markets: replicationmarkets.com

On Replication Markets, volunteer forecasters try to predict whether a given study's results will be replicated with high power. Rewards are monetary, but only given out to the top few forecasters, and markets suffer from sometimes being dull.

The first week of each round is a survey round, which has some aspects of a Keynesian beauty contest, because it's the results of the second round, not the ground truth, what is being forecasted. This second round then tries to predict what would happen if the studies were in fact subject to a replication, which a select number of studies then undergo.

There is a part of me which dislikes this setup: here was I, during the first round, forecasting to the best of my ability, when I realize that in some cases, I'm going to improve the aggregate and be punished for this, particularly when I have information which I expect other market participants to not have.

At first I thought that, cunningly, the results of the first round would be used as priors for the second round, but a [programming mistake](#) by the organizers revealed that they use a simple algorithm: claims with $p < .001$ start with a prior of 80%, $p < .01$ starts at 40%, and $p < .05$ starts at 30%.

In The News.

Articles and announcements in more or less traditional news media.

- [Locust-tracking application for the UN](#) (see [here](#) for a take by the Washington Post), using software originally intended to track the movements of air pollution. NOAA also sounds like a valuable organization: "NOAA Research enables better forecasts, earlier warnings for natural disasters, and a greater understanding of the Earth. Our role is to provide unbiased science to better manage the environment, nationally, and globally."
- [United Nations: World Economic Situation and Prospects as of mid-2020](#). A recent report is out, which predicts a 3.2% contraction of the global economy. Between 34 and 160 million people are expected to fall below the extreme poverty line this year. Compare with [Fitch ratings](#), which foresee a 4.6% decline in global GDP.
- [Fox News](#) and [Business Insider](#) report about the CDC forecasting 100k deaths by June the 1st, differently.
- Some transient content on 538 about [Biden vs past democratic nominees](#), about [Trump vs Biden polls](#) and about [the USA vicepresidential draft](#), and an old [review of the impact of VP candidates in USA elections](#) which seems to have aged well. 538 also brings us this overview of [models with unrealistic-yet-clearly-stated assumptions](#)
- [Why Economic Forecasting Is So Difficult in the Pandemic](#). Harvard Review Economists share their difficulties. Problems include "not knowing for sure what is going to happen", the government passing legislation uncharacteristically fast, sampling errors and reduced response rates from surveys, and lack of knowledge about epidemiology.
- [IBM releases new AI forecasting tool](#): "IBM Planning Analytics is an AI-infused integrated planning solution that automates planning, forecasting and budgeting." See [here](#) or [here](#) for a news take.
- Yahoo has automated finance forecast reporting. It took me a while (two months) to notice that the low-quality finance articles that were popping up in my google alerts were machine-generated. See [Synovus Financial Corp. Earnings Missed Analyst Estimates: Here's What Analysts Are Forecasting Now](#), [Wienerberger AG Earnings Missed Analyst Estimates: Here's What Analysts Are Forecasting Now](#), [Park Lawn Corporation Earnings Missed Analyst Estimates: Here's What Analysts Are Forecasting Now](#); they have a similar structure, paragraph per paragraph, and seem to have been generated from a template which changes a little bit depending on the data (they seem to have different templates for very positive, positive, neutral and negative change). To be clear, I could program something like this given a good finance api and a spare week/month, and in fact did so a couple of years ago for an automatic poetry generator, *but I didn't notice because I wasn't paying attention.*

- [Wimbledon organisers set to net £100 million insurance payout after taking out infectious diseases cover following 2003 SARS outbreak, with tournament now cancelled because of coronavirus](#). Cheers to Wimbledon.
- [The Post ranks the top 10 faces in New York sports today](#), accompanied by [Pitfall to forecasting top 10 faces of New York sports right now](#). Comparison with the historical situation: Check. Considering alternative hypothesis: Check. Communicating uncertainty to the reader in an effective manner: Check. Putting your predictions out to be judged: Check.
- [In Forecasting Hurricane Dorian, Models Fell Short](#) (and see [here](#) for the National Hurricane Center report). "Hurricane forecasters and the models they depend on failed to anticipate the strength and impact of last year's deadliest storm." On the topic of weather, see also [Nowcasting the weather in Africa](#) to reduce fatalities, and [Misunderstanding Of Coronavirus Predictions Is Eerily Similar To Weather Forecasting](#), Forbes speculates.
- [Pan-African Heatwave Health Hazard Forecasting](#). "The main aim, is to raise the profile of heatwaves as a hazard on a global scale. Hopefully, the project will add evidence to this sparse research area. It could also provide the basis for a heat early warning system." The project looks to be in its early stages, yet nonetheless interesting.
- [Nounós Creamery uses demand-forecasting platform to improve production process](#). The piece is shameless advertising, but it's still an example of predictive models used out in the wild in industry.

Grab Bag

Podcasts, blogposts, papers, tweets and other recent nontraditional media.

- Some interesting discussion about forecasting over at Twitter, in [David Manheim](#)'s and [Philip Tetlock](#)'s accounts, some of which have been incorporated into this newsletter. [This twitter thread](#) contains some discussion about how Good Judgement Open, Metaculus and expert forecasters fare against each other, but note the caveats by @LinchZhang: "For Survey 10, Metaculus said that question resolution was on 4pm ET Sunday, a lot of predictors (correctly) gauged that the data update on Sunday will be delayed and answered the letter rather than the spirit of the question (Metaculus ended up resolving it ambiguous)." [This thread](#) by Marc Lipsitch has become popular, and I personally also enjoyed [these two](#) twitter threads by Linchuan Zhang, on forecasting mistakes.
- [SlateStarCodex](#) brings us a hundred more predictions for 2020. Some analysis by Zvi Mowshowitz [here](#) and by user [Bucky](#).
- [FLI Podcast: On Superforecasting with Robert de Neufville](#). I would have liked to see a more intense drilling on some of the points. It references [The NonProphets Podcast](#), which looks like it has some more in-depth stuff. Some quotes:

So it's not clear to me that our forecasts are necessarily affecting policy. Although it's the kind of thing that gets written up in the news and who knows how much that affects people's opinions, or they talk about it at Davos and maybe those people go back and they change what they're doing.

I wish it were used better. If I were the advisor to a president, I would say you should create a predictive intelligence unit using superforecasters. Maybe give them access to some classified information, but even using open source information, have them predict probabilities of certain kinds of things and then develop a system for using that in your decision making. But I think we're a fair ways away from that. I don't know any interest in that in the current administration.

Now one thing I think is interesting is that often people, they're not interested in my saying, "There's a 78% chance of something happening." What they want to know is, how did I get there? What is my arguments? That's not unreasonable. I really like

thinking in terms of probabilities, but I think it often helps people understand what the mechanism is because it tells them something about the world that might help them make a decision. So I think one thing that maybe can be done is not to treat it as a black box probability, but to have some kind of algorithmic transparency about our thinking because that actually helps people, might be more useful in terms of making decisions than just a number.

- [Space Weather Challenge and Forecasting Implications of Rossby Waves](#). Recent advances may help predict solar flares better. I don't know how bad the worst solar flare could be, and how much a two year warning could buy us, but I tend to view developments like this very positively.
- [An analogy-based method for strong convection forecasts in China using GFS forecast data](#). "Times in the past when the forecast parameters are most similar to those forecast at the current time are identified by searching a large historical numerical dataset", and this is used to better predict one particular class of meteorological phenomena. See [here](#) for a press release.
- The Cato Institute releases [12 New Immigration Ideas for the 21st Century](#), including two from Robin Hanson: Choosing Immigrants through Prediction Markets & Transferable Citizenship. The first idea is to have prediction markets forecast the monetary value of taking in immigrants, and decide accordingly, then rewarding forecasters according to their accuracy in predicting e.g. how much said immigrants pay in taxes.
- [A General Approach for Predicting the Behavior of the Supreme Court of the United States](#). What seems to be a pretty simple algorithm (a random forest!) seems to do pretty well (70% accuracy). Their feature set is rich but doesn't seem to include ideology. It was written in 2017; today, I'd expect that a random bright highschooler might be able to do much better.
- [From Self-Prediction to Self-Defeat: Behavioral Forecasting, Self-Fulfilling Prophecies, and the Effect of Competitive Expectations](#). Abstract: Four studies explored behavioral forecasting and the effect of competitive expectations in the context of negotiations. Study 1 examined negotiators' forecasts of how they would behave when faced with a very competitive versus a less competitive opponent and found that negotiators believed they would become more competitive. Studies 2 and 3 examined actual behaviors during negotiation and found that negotiators who expected a very competitive opponent actually became less competitive, as evidenced by setting lower, less aggressive reservation prices, making less demanding counteroffers, and ultimately agreeing to lower negotiated outcomes. Finally, Study 4 provided a direct test of the disconnection between negotiators' forecasts for their behavior and their actual behaviors within the same sample and found systematic errors in behavioral forecasting as well as evidence for the self-fulfilling effects of possessing a competitive expectation.
- [Neuroimaging results altered by varying analysis pipelines](#). Relevant paragraph: "the authors ran separate 'prediction markets', one for the analysis teams and one for researchers who did not participate in the analysis. In them, researchers attempted to predict the outcomes of the scientific analyses and received monetary payouts on the basis of how well they predicted performance. Participants — even researchers who had direct knowledge of the data set — consistently overestimated the likelihood of significant findings". Those who had more knowledge did slightly better, however.
- [Forecasting s-curves is hard](#): Some clear visualizations of what it says on the title.
- [Forecasting state expenses for budget is always a best guess](#); exactly what it says on the tin. Problem could be solved with a prediction market or forecasting tournament.
- [Fashion Trend Forecasting](#) using Instagram and baking preexisting knowledge into NNs.

- [The advantages and limitations of forecasting](#). A short and sweet blog post, with a couple of forecasting anecdotes and zingers.

Negative examples.

I have found negative examples to be useful as a mirror with which to reflect on my own mistakes; highlighting them may also be useful for shaping social norms. [Andrew Gelman](#) continues to fast-pacedly produce blogposts on this topic. Meanwhile, amongst mortals:

- [Kelsey Piper of Vox harshly criticizes the IHME model](#). "Some of the factors that make the IHME model unreliable at predicting the virus may have gotten people to pay attention to it;" or "Other researchers found the true deaths were outside of the 95 percent confidence interval given by the model 70 percent of the time."
- The [Washington post](#) offers a highly partisan view of Trump's chances of winning the election. The author, having already made a past prediction, and seeing as how other media outlets offer a conflicting perspective, rejects the information he's learnt, and instead can only come up with more reasons which confirm his initial position. Problem could be solved with a prediction market or forecasting tournament.
- [California politics pretends to be about recession forecasts](#). See also: [Simulacra levels](#); the article is at least three levels removed from consideration about bare reality. Key quote, about a given forecasting model: "It's just preposterously negative... How can you say that out loud without giggling?" See also some more prediction ping-pong, this time in New Jersey, [here](#). Problem could be solved with a prediction market or forecasting tournament.
- [What Is the Stock Market Even for Anymore?](#). A New York Times claims to have predicted that the market was going to fall (but can't prove it with, for example, a tweet, or a hash of a tweet), and nonetheless lost significant amounts of his own funds. ("The market dropped another 1,338 points the next day, and though my funds were tanking along with almost everyone else's, I found some empty satisfaction, at least, in my prognosticating.") The rest of the article is about said reported being personally affronted with the market not falling further ("the stock market's shocking resilience (at least so far) has looked an awful lot like indifference to the Covid-19 crisis and the economic calamity it has brought about. The optics, as they say, are terrible.")
- [Forecasting drug utilization and expenditure: ten years of experience in Stockholm](#). A normally pretty good forecasting model had the bad luck of not foreseeing a Black Swan, and sending a study to a journal just before a pandemic, so that it's being published now. They write: "According to the forecasts, the total pharmaceutical expenditure was estimated to increase between 2 and 8% annually. Our analyses showed that the accuracy of these forecasts varied over the years with a mean absolute error of 1.9 percentage points." They further conclude: "Based on the analyses of all forecasting reports produced since the model was established in Stockholm in the late 2000s, we demonstrated that it is feasible to forecast pharmaceutical expenditure with a reasonable accuracy." Presumably, this has increased further because of covid, sending the mean absolute error through the roof. If the author of this paper bites you, you become a Nassim Taleb.
- Some films are so bad it's funny. [This article fills the same niche](#) for forecasting. It has it all: Pythagorean laws of vibration, epicycles, an old and legendary master with mystical abilities, 90 year predictions which come true. Further, from the [Wikipedia entry](#): "He told me that his famous father could not support his family by trading but earned his living by writing and selling instructional courses."
- [Austin Health Official Recommends Cancelling All 2020 Large Events, Despite Unclear Forecasting](#). Texan article does not consider the perspective that one might want to cancel large events precisely *because* of the forecasting uncertainty.
- [Auditor urges more oversight, better forecasting at the United State's Department of Transport](#): "Instead of basing its spending plan on project-specific cost estimates, Wood said, the agency uses prior-year spending. That forecasting method doesn't account for cost increases or for years when there are more projects in the works." The budget of the

organization is \$5.9 billion. Problem could be solved with a prediction market or forecasting tournament.

Long content

This section contains items which have recently come to my attention, but which I think might still be relevant not just this month, but throughout the years. Content in this section may not have been published in the last month.

- [How to evaluate 50% predictions](#). "I commonly hear (sometimes from very smart people) that 50% predictions are meaningless. I think that this is wrong."
- [Named Distributions as Artifacts](#). On how the named distributions we use (the normal distribution, etc.), were selected for being easy to use in pre-computer eras, rather than on being a good ur-prior on distributions for phenomena in this universe.
- [The fallacy of placing confidence in confidence intervals](#). On how the folk interpretation of confidence intervals can be misguided, as it conflates: a. the long-run probability, before seeing some data, that a procedure will produce an interval which contains the true value, and b. and the probability that a particular interval contains the true value, after seeing the data. This is in contrast to Bayesian theory, which can use the information in the data to determine what is reasonable to believe, in light of the model assumptions and prior information. I found their example where different confidence procedures produce 50% confidence intervals which are nested inside each other particularly funny. Some quotes:

Using the theory of confidence intervals and the support of two examples, we have shown that CIs do not have the properties that are often claimed on their behalf. Confidence interval theory was developed to solve a very constrained problem: how can one construct a procedure that produces intervals containing the true parameter a fixed proportion of the time? Claims that confidence intervals yield an index of precision, that the values within them are plausible, and that the confidence coefficient can be read as a measure of certainty that the interval contains the true value, are all fallacies and unjustified by confidence interval theory.

"I am not at all sure that the 'confidence' is not a 'confidence trick.' Does it really lead us towards what we need – the chance that in the universe which we are sampling the parameter is within these certain limits? I think it does not. I think we are in the position of knowing that either an improbable event has occurred or the parameter in the population is within the limits. To balance these things we must make an estimate and form a judgment as to the likelihood of the parameter in the universe that is, a prior probability – the very thing that is supposed to be eliminated."

The existence of multiple, contradictory long-run probabilities brings back into focus the confusion between what we know before the experiment with what we know after the experiment. For any of these confidence procedures, we know before the experiment that 50 % of future CIs will contain the true value. After observing the results, conditioning on a known property of the data — such as, in this case, the variance of the bubbles — can radically alter our assessment of the probability.

"You keep using that word. I do not think it means what you think it means." Inigo Montoya, The Princess Bride (1987)

- [Psychology of Intelligence Analysis](#), courtesy of the American Central Intelligence Agency, seemed interesting, and I read chapters 4, 5 and 14. Sometimes forecasting looks like reinventing intelligence analysis; from that perspective, I've found this reference work useful. Thanks to EA Discord user @Willow for bringing this work to my attention.

- Chapter 4: Strategies for Analytical Judgement. Discusses and compares the strengths and weaknesses of four tactics: situational analysis (inside view), applying theory, comparison with historical situations, and immersing oneself on the data. It then brings up several suboptimal tactics for choosing among hypotheses.
- Chapter 5: When does one need more information, and in what shapes does new information come from?

Once an experienced analyst has the minimum information necessary to make an informed judgment, obtaining additional information generally does not improve the accuracy of his or her estimates. Additional information does, however, lead the analyst to become more confident in the judgment, to the point of overconfidence.

Experienced analysts have an imperfect understanding of what information they actually use in making judgments. They are unaware of the extent to which their judgments are determined by a few dominant factors, rather than by the systematic integration of all available information. Analysts actually use much less of the available information than they think they do.

There is strong experimental evidence, however, that such self-insight is usually faulty. The expert perceives his or her own judgmental process, including the number of different kinds of information taken into account, as being considerably more complex than is in fact the case. Experts overestimate the importance of factors that have only a minor impact on their judgment and underestimate the extent to which their decisions are based on a few major variables. In short, people's mental models are simpler than they think, and the analyst is typically unaware not only of which variables should have the greatest influence, but also which variables actually are having the greatest influence.

- Chapter 14: A Checklist for Analysts. "Traditionally, analysts at all levels devote little attention to improving how they think. To penetrate the heart and soul of the problem of improving analysis, it is necessary to better understand, influence, and guide the mental processes of analysts themselves." The Chapter also contains an Intelligence Analysis reading list.
- [The Limits of Prediction: An Analyst's Reflections on Forecasting](#), also courtesy of the American Central Intelligence Agency. On how intelligence analysts should inform their users of what they are and aren't capable of. It has some interesting tidbits and references on predicting discontinuities. It also suggests some guiding questions that the analyst may try to answer for the policymaker.
 - What is the context and reality of the problem I am facing?
 - How does including information on new developments affect my problem/issue?
 - What are the ways this situation could play out?
 - How do we get from here to there? and/or What should I be looking out for?
- "We do not claim our assessments are infallible. Instead, we assert that we offer our most deeply and objectively based and carefully considered estimates."
- [How to Measure Anything](#), a review. "Anything can be measured. If a thing can be observed in any way at all, it lends itself to some type of measurement method. No matter how "fuzzy" the measurement is, it's still a measurement if it tells you more than you knew before. And those very things most likely to be seen as immeasurable are, virtually always, solved by relatively simple measurement methods."
- The World Meteorological organization, on their mandate to guarantee that [no one is surprised by a flood](#). Browsing the webpage it seems that the organization is either a Key

Organization Safeguarding the Vital Interests of the World or Just Another of the Many Bureaucracies Already in Existence, but it's unclear to me how to differentiate between the two. One clue may be their recent [Caribbean workshop on impact-based forecasting and risk scenario planning](#), with the narratively unexpected and therefore salient presence of Gender Bureaus.

- [95%-ile isn't that good](#): "Reaching 95%-ile isn't very impressive because it's not that hard to do."
- [The Backwards Arrow of Time of the Coherently Bayesian Statistical Mechanic](#): Identifying thermodynamic entropy with the Bayesian uncertainty of an ideal observer leads to problems, because as the observer observes more about the system, they update on this information, which in expectation reduces uncertainty, and thus entropy. But entropy increases with time.
 - This might be interesting to students in the tradition of E.T. Jaynes: for example, the paper directly conflicts with this LessWrong post: [The Second Law of Thermodynamics, and Engines of Cognition](#), part of *Rationality, From AI to Zombies*. The way out might be to postulate that actually, the Bayesian updating process itself would increase entropy, in the form of e.g., the work needed to update bits on a computer. Any applications to Christian lore are left as an exercise for the reader. Otherwise, seeing two bright people being cogently convinced of different perspectives does something funny to my probabilities: it pushes them towards 50%, but also increases the expected time I'd have to spend on the topic to move them away from 50%.
- [Behavioral Problems of Adhering to a Decision Policy](#)

Our judges in this study were eight individuals, carefully selected for their expertise as handicappers. Each judge was presented with a list of 88 variables culled from the past performance charts. He was asked to indicate which five variables out of the 88 he would wish to use when handicapping a race, if all he could have was five variables. He was then asked to indicate which 10, which 20, and which 40 he would use if 10, 20, or 40 were available to him.

We see that accuracy was as good with five variables as it was with 10, 20, or 40. The flat curve is an average over eight subjects and is somewhat misleading. Three of the eight actually showed a decrease in accuracy with more information, two improved, and three stayed about the same. All of the handicappers became more

confident in their judgments as information increased.

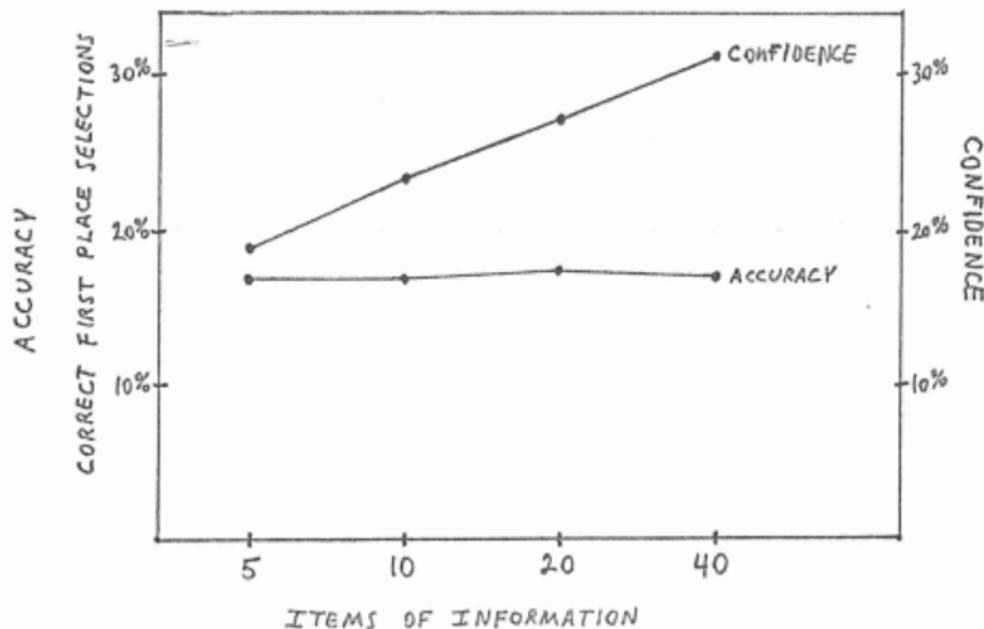


Figure 4. Mean changes in confidence and accuracy with increasing amounts of information.

- The study contains other nuggets, such as:
- An experiment on trying to predict the outcome of a given equation. When the feedback has a margin of error, this confuses respondents.
- "However, the results indicated that subjects often chose one gamble, yet stated a higher selling price for the other gamble"
- "We figured that a comparison between two students along the same dimension should be easier, cognitively, than a 13 comparison between different dimensions, and this ease of use should lead to greater reliance on the common dimension. The data strongly confirmed this hypothesis. Dimensions were weighted more heavily when common than when they were unique attributes. Interrogation of the subjects after the experiment indicated that most did not wish to change their policies by giving more weight to common dimensions and they were unaware that they had done so."
- "The message in these experiments is that the amalgamation of different types of information and different types of values into an overall judgment is a difficult cognitive process. In our attempts to ease the strain of processing information, we often resort to judgmental strategies that do an injustice to the underlying values and policies that we're trying to implement."
- "A major problem that a decision maker faces in his attempt to be faithful to his policy is the fact that his insight into his own behavior may be inaccurate. He may not be aware of the fact that he is employing a different policy than he thinks he's using. This problem is illustrated by a study that Dan Fleissner, Scott Bauman, and I did, in which 13 stockbrokers and five graduate students served as subjects. Each subject evaluated the potential capital appreciation of 64 securities. [...] A mathematical model was then constructed to predict each subject's judgments. One output from the model was an index of the relative importance of each of the eight information items in determining each subject's judgments [...] Examination of Table 4 shows that the broker's perceived weights did not relate closely to the weights derived from their actual judgments.

- I informally replicated this.
- As remedies they suggest to create a model by eliciting the expert, either by having the expert make a large number of judgments and distilling a model, or by asking the expert what they think the most important factors are. A third alternative suggested is computer assistance, so that the experiment participants become aware of which factors influence their judgment.

- [Immanuel Kant, on Betting](#)

Vale.

Conflicts of interest: Marked as (c.o.i) throughout the text.

Note to the future: All links are automatically added to the Internet Archive. In case of link rot, go [there](#).

Forecasting Newsletter. June 2020.

Highlights

1. Facebook launches [Forecast](#), a community for crowdsourced predictions.
2. Foretell, a forecasting tournament by the Center for Security and Emerging Technology, is now [open](#).
3. [A Preliminary Look at Metaculus and Expert Forecasts](#): Metaculus forecasters do better.

Sign up [here](#), view this newsletter on the EA Forum [here](#), or browse past newsletters [here](#)

Index

- Highlights.
- In the News.
- Prediction Markets & Forecasting Platforms.
- Negative Examples.
- Hard to Categorize.
- Long Content.

In the News.

- Facebook releases a forecasting app ([link to the app](#), [press release](#), [TechCrunch take](#), [hot-takes](#)). The release comes before Augur v2 launches, and it is easy to speculate that it might end up being combined with Facebook's stablecoin, Libra.
- The Economist has a new electoral model out ([article](#), [model](#)) which gives Trump an 11% chance of winning reelection. Given that Andrew Gelman was involved, I'm hesitant to criticize it, but it seems a tad overconfident. See [here](#) for Gelman addressing objections similar to my own.
- [COVID-19 vaccine before US election](#). Analysts see White House pushing through vaccine approval to bolster Trump's chances of reelection before voters head to polls. "All the datapoints we've collected make me think we're going to get a vaccine prior to the election," Jared Holz, a health-care strategist with Jefferies, said in a phone interview. The current administration is "incredibly incentivized to approve at least one of these vaccines before Nov. 3."
- "[Israeli Central Bank Forecasting Gets Real During Pandemic](#)". Israeli Central Bank is using data to which it has real-time access, like credit-card spending, instead of lagging indicators.
- [Google](#) produces wind schedules for wind farms. "The result has been a 20 percent increase in revenue for wind farms". See [here](#) for essentially the same thing on solar forecasting.

- Survey of macroeconomic researchers predicts economic recovery will take years, reports [538](#).

Prediction Markets & Forecasting platforms.

Ordered in subjective order of importance:

- Foretell, a forecasting tournament by the Center for Security and Emerging Technology, is now [open](#). I find the thought heartening that this might end up influencing bona-fide politicians.
- Metaculus
 - posted [A Preliminary Look at Metaculus and Expert Forecasts](#): Metaculus forecasters do better, and the piece is a nice reference point.
 - was featured in [Forbes](#).
 - announced their [Metaculus Summer Academy](#): "an introduction to forecasting for those who are relatively new to the activity and are looking for a fresh intellectual pursuit this summer"
- [Replication Markets](#) might add a new round with social and behavioral science claims related to COVID-19, and a preprint market, which would ask participants to forecast items like publication or citation. Replication Markets is also asking for more participants, with the catchline "If they are knowledgeable and opinionated, Replication Markets is the place to be to make your opinions really count."
- Good Judgement family
 - [Good Judgement Open](#): Superforecasters were [able](#) to detect that Russia and the USA would in fact undertake some (albeit limited) form of negotiation, and do so much earlier than the general public, even while posting their reasons in full view.
 - Good Judgement Analytics continues to provide its [COVID-19 dashboard](#).
- [PredictIt](#) & [Election Betting Odds](#). I stumbled upon an old 538 piece on fake polls: [Fake Polls are a Real Problem](#). Some polls may have been conducted by PredictIt traders in order to mislead or troll other PredictIt traders; all in all, an amusing example of how prediction markets could encourage worse information.
- [An online prediction market with reputation points](#), implementing an [idea](#) by Paul Christiano. As of yet slow to load.
- Augur:
 - [An overview of the platform and of v2 modifications](#).
 - Augur also happens to have a [blog](#) with some interesting tidbits, such as the extremely clickbaity [How One Trader Turned \\$400 into \\$400k with Political Futures](#) ("I find high volume markets...like the Democratic Nominee market or the 2020 Presidential Winner market... and what I'm doing is I'm just getting in line at the 'buy' price and waiting my turn until my orders get filled. Then when those orders get filled I just sell them for 1c more.")

- [Coronavirus Information Markets](#) is down to ca. \$12000 in trading volume; it seems like they didn't take off.

Negative examples.

- World powers to converge on strategies for presenting COVID-19 information to make forecasters' jobs more interesting:
 - [Brazil stops releasing COVID-19 death toll and wipes data from official site](#).
 - Meanwhile, in Russia, [St Petersburg issues 1,552 more death certificates in May than last year, but Covid-19 toll was 171](#).
 - In the US, [CDC wants states to count 'probable' coronavirus cases and deaths, but most aren't doing it](#)
 - [India has the fourth-highest number of COVID-19 cases, but the Government denies community transmission](#)
 - One suspects that this denial is political, because India is otherwise [being extremely competent](#) in weather forecasting.
- Youyang Gu's model, widely acclaimed as one of the best coronavirus models for the US, produces 95% confidence intervals which [seem too narrow](#) when extended to [Pakistan](#).
- Some discussion on [twitter](#): "Only a fool would put a probability on whether the EU and the UK will agree a trade deal", says Financial Times correspondent, and other examples.

Hard to categorize.

- [A Personal COVID-19 Postmortem](#), by FHI researcher [David Manheim](#).

I think it's important to clearly and publicly admit when we were wrong. It's even better to diagnose why, and take steps to prevent doing so again.

COVID-19 is far from over, but given my early stance on a number of questions regarding COVID-19, this is my attempt at a public personal review to see where I was wrong.

- [FantasyScotus](#) beat [GoodJudgementOpen](#) on legal decisions. I'm still waiting to see whether [Hollywood Stock Exchange](#) will also beat GJOpen on [film predictions](#).
- [How does pandemic forecasting resemble the early days of weather forecasting](#); what lessons can the USA learn from the later about the former? An example would be to create an organization akin to the National Weather Center, but for forecasting.
- Linch Zhang, a COVID-19 forecaster with an excellent track-record, is doing an [Ask Me Anything](#), starting on Sunday the 7th; questions are welcome!
- [The Rules To Being A Sellside Economist](#). A fun read.

5. How to get attention: If you want to get famous for making big non-consensus calls, without the danger of looking like a muppet, you should adopt 'the 40% rule'. Basically you can forecast whatever you want with a probability of 40%. Greece to quit the euro? Maybe! Trump to fire Powell and hire his daughter as the new Fed chair? Never say never! 40% means the odds will be greater than anyone else is saying, which is why your clients need to listen to your warning, but also that they shouldn't be too surprised if, you know, the extreme event doesn't actually happen.

- [How to improve space weather forecasting](#) (see [here](#) for the original paper):

For instance, the National Oceanic and Atmospheric Administration's Deep Space Climate Observatory (DSCOVR) satellite sits at the location in space called L1, where the gravitational pulls of Earth and the Sun cancel out. At this point, which is roughly 1.5 million kilometers from Earth, or barely 1% of the way to the Sun, detectors can provide warnings with only short lead times: about 30 minutes before a storm hits Earth in most cases or as little as 17 minutes in advance of extremely fast solar storms.

- [Coup cast](#): A site that estimates the yearly probability of a coup. The color coding is misleading; click on the countries instead.
- [Prediction = Compression](#). "Whenever you have a prediction algorithm, you can also get a correspondingly good compression algorithm for data you already have, and vice versa."
 - Other LessWrong posts which caught my attention were [Betting with Mandatory Post-Mortem](#) and [Radical Probabilism](#)
- [Box Office Pro](#) looks at some factors around box-office forecasting.

Long Content.

- [When the crowds aren't wise](#); a sober overview, with judicious use of [Cordocet's jury theorem](#)

Suppose that each individual in a group is more likely to be wrong than right because relatively few people in the group have access to accurate information. In that case, the likelihood that the group's majority will decide correctly falls toward zero as the size of the group increases.

Some prediction markets fail for just this reason. They have done really badly in predicting President Bush's appointments to the Supreme Court, for example. Until roughly two hours before the official announcement, the markets were essentially ignorant of the existence of John Roberts, now the chief justice of the United States. At the close of a prominent market just one day before his nomination, "shares" in Judge Roberts were trading at \$0.19—representing an estimate that Roberts had a 1.9% chance of being nominated.

Why was the crowd so unwise? Because it had little accurate information to go on; these investors, even en masse, knew almost nothing about the internal deliberations in the Bush administration. For similar reasons,

prediction markets were quite wrong in forecasting that weapons of mass destruction would be found in Iraq and that special prosecutor Patrick Fitzgerald would indict Deputy Chief of Staff Karl Rove in late 2005.

- [A review of Tetlock's 'Superforecasting' \(2015\)](#), by Dominic Cummings. Cummings then went on to hire one such superforecaster, which then resigned over a [culture war](#) scandal, characterized by adversarial selection of quotes which indeed are outside the British Overton Window. Notably, Dominic Cummings then told reporters to "Read Philip Tetlock's *Superforecasters*, instead of political pundits who don't know what they're talking about."
- [Assessing the Performance of Real-Time Epidemic Forecasts: A Case Study of Ebola in the Western Area Region of Sierra Leone, 2014-15](#). The one caveat is that their data is much better than coronavirus data, because Ebola symptoms are more evident; otherwise, pretty interesting:

Real-time forecasts based on mathematical models can inform critical decision-making during infectious disease outbreaks. Yet, epidemic forecasts are rarely evaluated during or after the event, and there is little guidance on the best metrics for assessment.

...good probabilistic calibration was achievable at short time horizons of one or two weeks ahead but model predictions were increasingly unreliable at longer forecasting horizons.

This suggests that forecasts may have been of good enough quality to inform decision making based on predictions a few weeks ahead of time but not longer, reflecting the high level of uncertainty in the processes driving the trajectory of the epidemic.

Comparing different versions of our model to simpler models, we further found that it would have been possible to determine the model that was most reliable at making forecasts from early on in the epidemic. This suggests that there is value in assessing forecasts, and that it should be possible to improve forecasts by checking how good they are during an ongoing epidemic.

One forecast that gained particular attention during the epidemic was published in the summer of 2014, projecting that by early 2015 there might be 1.4 million cases. This number was based on unmitigated growth in the absence of further intervention and proved a gross overestimate, yet it was later highlighted as a "call to arms" that served to trigger the international response that helped avoid the worst-case scenario.

Methods to assess probabilistic forecasts are now being used in other fields, but are not commonly applied in infectious disease epidemiology

The deterministic SEIR model we used as a null model performed poorly on all forecasting scores, and failed to capture the downturn of the epidemic in Western Area.

On the other hand, a well-calibrated mechanistic model that accounts for all relevant dynamic factors and external influences could, in principle, have been used to predict the behaviour of the epidemic reliably and precisely. Yet, lack of detailed data on transmission routes and risk factors precluded

the parameterisation of such a model and are likely to do so again in future epidemics in resource-poor settings.

- In the selection of quotes above, we gave an example of a forecast which ended up overestimating the incidence, yet might have "served as a call to arms". It's maybe a real-life example of a forecast changing the true result, leading to a fixed point problem, like the ones hypothesized in the parable of the [Predict-O-Matic](#).
 - It would be a fixed point problem if [forecast above the alarm threshold] → epidemic being contained, but [forecast below the alarm threshold] → epidemic not being contained.
 - Maybe the fix-point solution, i.e., the most self-fulfilling (and thus, accurate) forecast, would have been a forecast on the edge of the alarm threshold, which would have ended up leading to mediocre containment.
 - The [troll polls](#) created by PredictIt traders are perhaps a more clear cut example of Predict-O-Matic problems.
- [Calibration Scoring Rules for Practical Prediction Training](#). I found it most interesting when considering how Brier and log rules didn't have all the pedagogic desiderata.
 - I also found the following derivation of the logarithmic scoring rule interesting. Consider: If you assign a probability to n events, then the combined probability of these events is $p_1 \times p_2 \times p_3 \times \dots \times p_n$. Taking logarithms, this is $\log(p_1 \times p_2 \times p_3 \times \dots \times p_n) = \sum \log(p_n)$, i.e., the logarithmic scoring rule.
- [Binary Scoring Rules that Incentivize Precision](#). The results (the closed-form of scoring rules which minimize a given forecasting error) are interesting, but the journey to get there is kind of a drag, and ultimately the logarithmic scoring rule ends up being pretty decent according to their measure of error.
 - Opinion: I'm not sure whether their results are going to be useful for things I'm interested in (like human forecasting tournaments, rather than Kaggle data analysis competitions). In practice, what I might do if I wanted to incentivize precision is to ask myself if this is a question where the answer is going to be closer to 50%, or closer to either of 0% or 100%, and then use either the Brier or the logarithmic scoring rules. That is, I don't want to minimize an L-norm of the error over [0,1], I want to minimize an L-norm over the region I think the answer is going to be in, and the paper falls short of addressing that.
- [How Innovation Works—A Review](#). The following quote stood out for me:

Ridley points out that there have always been opponents of innovation. Such people often have an interest in maintaining the status quo but justify their objections with reference to the precautionary principle.

- [A list of prediction markets](#), and their fates, maintained by Jacob Lagerros. Like most startups, most prediction markets fail.

Note to the future: All links are added automatically to the Internet Archive. In case of link rot, go [here](#)

"I beseech you, in the bowels of Christ, think it possible that you may be mistaken." [Oliver Cromwell](#)

Forecasting Newsletter: July 2020.

Highlights

- Social Science Prediction Platform [launches](#).
- Ioannidis and Taleb [discuss](#) optimal response to COVID-19.
- Report tries to [foresee](#) the (potentially quite high) dividends of conflict prevention from 2020 to 2030.

Index

- Highlights.
- Prediction Markets & Forecasting Platforms.
- New undertakings.
- Negative Examples.
- News & Hard to Categorize Content.
- Long Content.

Sign up [here](#), browse past newsletters [here](#), or view it on the EA forum [here](#).

Prediction Markets & Forecasting Platforms.

Ordered in subjective order of importance:

- Metaculus continues hosting great discussion.
- In particular, it has recently hosted some high-quality [AI questions](#).
- User @alexrlj, a moderator on the platform, [offers on the EA forum](#) to operationalize questions and post them on Metaculus, for free. This hasn't been picked up by the EA Forum algorithms, but the offer seems to me to be quite valuable. Some examples of things you might want to see operationalized and forecasted: the funding your organization will receive in 2020, whether any particularly key bills will become law, whether GiveWell will change their top charities, etc.
- [Foretell](#) is a prediction market by the University of Georgetown's Center for Security and Emerging Technology, focused on questions relevant to technology-security policy, and on bringing those forecasts to policy-makers.
- Some EAs, such as myself or a mysterious user named *foretold*, feature on the top spots of their (admittedly quite young) leaderboard.
- I also have the opportunity to create a team on the site: if you have a proven track record and would be interested in joining such a team, get in touch before the 10th

of August.

- [Replication Markets](#)
- published their [first paper](#)
- had some difficulties with cheaters:

"The Team at Replication Markets is delaying announcing the Round 8 Survey winners because of an investigation into coordinated forecasting among a group of participants. As a result, eleven accounts have been suspended and their data has been excluded from the study. Scores are being recalculated and prize announcements will go out soon."

- Because of how Replication Markets are structured, I'm betting the cheating was by manipulating the Keynesian beauty contest in a [Predict-O-Matic](#) fashion. That is, cheaters could have coordinated to output something surprising during the Keynesian Beauty Contest round, and then make that surprising thing come to happen during the market trading round. Charles Twardy, principal investigator at Replication Markets, gives a more positive take on the Keynesian beauty contest aspects of Replication Markets [here](#).
- still have Round 10 open until the 3rd of August.
- At the Good Judgement family, Good Judgement Analytics continues to provide its [COVID-19 dashboard](#).

Modeling is a very good way to explain how a virus will move through an unconstrained herd. But when you begin to put in constraints" — mask mandates, stay-at-home orders, social distancing — "and then the herd has agency whether they're going to comply, at that point, human forecasters who are very smart and have read through the models, that's where they really begin to add value. – Marc Koehler, Vice President of Good Judgement, Inc., in a [recent interview](#)

- [Highly Speculative Estimates](#), an interface, library and syntax to produce distributional probabilistic estimates led by Ozzie Gooen, now accepts functions as part of its input, such that more complicated inputs like the following are now possible:

```
# Variable: Number of ice creams an unsupervised child has consumed,  
# when left alone in an ice cream shop.  
  
# Current time (hours passed)  
t=10  
  
# Scenario with lots of uncertainty  
w_1 = 0.75 ## Weight for this scenario.  
min_uncertain(t) = t*2  
max_uncertain(t) = t*20
```

```

# Optimistic scenario
w_2 = 0.25 ## Weight for the optimistic scenario
min_optimistic(t) = 1*t
max_optimistic(t) = 3*t
mean(t) = (min_optimistic(t) + max_optimistic(t)/2)
stdev(t) = t*(2)^{1/2}

# Overall guess
## A long-tailed lognormal for the uncertain scenario
## and a tight normal for the optimistic scenario

mm(min_uncertain(t) to max_uncertain(t), normal(mean(t), stdev(t)), [w_1, w_2])

## Compare with: mm(2 to 20, normal(2, 1.4142), [0.75, 0.25])

```

- [PredictIt](#) & [Election Betting Odds](#) each give a 60%-ish to Biden.
- See [Limits of Current US Prediction Markets \(PredictIt Case Study\)](#), on how spread, transaction fees, withdrawal fees, interest rate which one could otherwise be earning, taxes, and betting limits make it so that:

"Current prediction markets are so bad in so many different ways that it simply is not surprising for people to know better than them, and it often is not possible for people to make money from knowing better."

- [Augur](#), a betting platform built on top of Ethereum, launches v2. Here are [two overviews](#) of the platform and of v2 modifications

New undertakings

- [Announcing the Launch](#) of the [Social Science Prediction Platform](#), a platform aimed at collecting and popularizing predictions of research results, in order to improve social science; see [this Science article](#) for the background motivation:

A new result builds on the consensus, or lack thereof, in an area and is often evaluated for how surprising, or not, it is. In turn, the novel result will lead to an updating of views. Yet we do not have a systematic procedure to capture the scientific views prior to a study, nor the updating that takes place afterward. What did people predict the study would find? How would knowing this result affect the prediction of findings of future, related studies?

A second benefit of collecting predictions is that they [...] can also potentially help to mitigate publication bias. However, if priors are collected before carrying out a study, the results can be compared to the average expert prediction, rather than to the null hypothesis of no effect. This would allow researchers to confirm that some results were unexpected, potentially making them more interesting and informative, because they indicate rejection of a prior held by the research community; this could contribute to alleviating publication bias against null results.

A third benefit of collecting predictions systematically is that it makes it possible to improve the accuracy of predictions. In turn, this may help with experimental design.

- On the one hand, I could imagine this having an impact, and the enthusiasm of the founders is contagious. On the other hand, as a forecaster I don't feel enticed by the platform: they offer a \$25 reward to grad students (which I am not), and don't spell it out for me why I would want to forecast on their platform as opposed to on [all the other alternatives available to me](#), even accounting for altruistic impact.
- [Ought](#) is a research lab building tools to delegate open-ended reasoning to AI & ML systems.
- Since concluding their initial factored cognition experiments in 2019, they've been building tools to capture and automate the reasoning process in forecasting: [Ergo](#), a library for integrating model-based and judgmental forecasting, and [Elicit](#), a tool built on top of Ergo to help forecasters express and share distributions.
- They've recently run small-scale tests exploring amplification and delegation of forecasting, such as: [Amplify Rohin's Prediction on AGI researchers & Safety Concerns](#), [Amplified forecasting: What will Buck's informed prediction of compute used in the largest ML training run before 2030 be?](#), and [Delegate a Forecast](#).
 - See also [Amplifying generalist research via forecasting](#), previous work in a similar direction which was also inspired by Paul Christiano's Iterated Distillation and Amplification agenda.
- In addition to studying factored cognition in the forecasting context, they are broadly interested in whether the EA community could benefit from better forecasting tools: they can be reached out to team@ought.org if you want to give them feedback or discuss their work.
- [The Pipeline Project](#) is a project similar to Replication Markets, by some of the same authors, to find out whether people can predict whether a given study will replicate. They offer authorship in an appendix, as well as a chance to get a token monetary compensation.
- [USAID's Intelligent Forecasting: A Competition to Model Future Contraceptive Use](#). "First, we will award up to 25,000 USD in prizes to innovators who develop an intelligent forecasting model—using the data we provide and methods such as artificial intelligence (AI)—to predict the consumption of contraceptives over three months. If implemented, the model should improve the availability of contraceptives and family planning supplies at health service delivery sites throughout a nationwide healthcare system. Second, we will award a Field Implementation Grant of approximately 100,000 to 200,000 USD to customize and test a high-performing intelligent forecasting model in Côte d'Ivoire."
- [Omen](#) is another cryptocurrency-based prediction market, which seems to use the same front-end (and probably back-end) as [Corona Information Markets](#). It's unclear what their advantages with respect to Augur are.

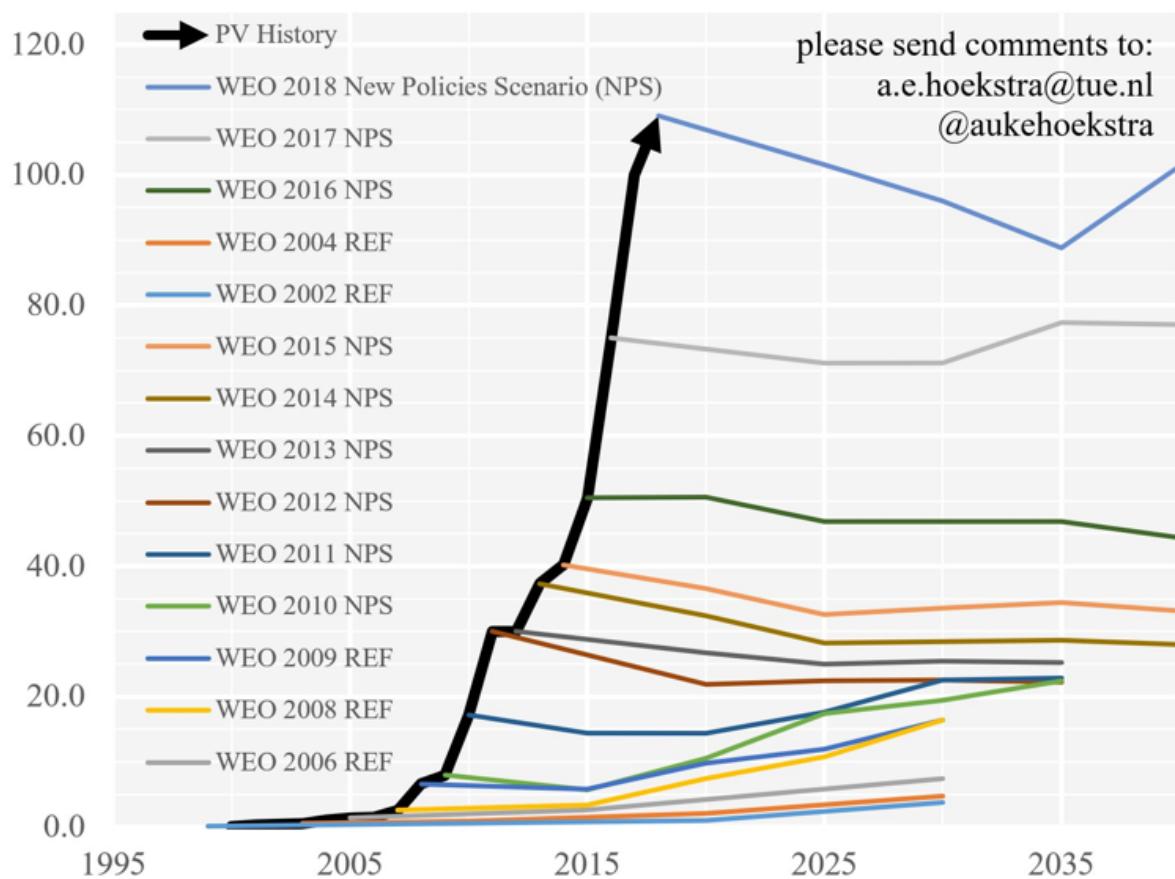
- [Yngve Høiseth](#) releases a prediction scorer, based on his previous work on Empiricast. In Python, but also available as a [REST API](#)

Negative Examples.

- The International Energy Agency had terrible forecasts on solar photo-voltaic energy production, until [recently](#):

Annual PV additions: historic data vs IEA WEO predictions

In GW of added capacity per year - source International Energy Agency - World Energy Outlook



...It's a scenario assuming current policies are kept and no new policies are added.

...the discrepancy basically implies that every year loads of unplanned subsidies are added... So it boils down to: it's not a forecast and any error you find must be attributed to that. And no you cannot see how the model works.

The IEA website explains the WEO process: "The detailed projections are generated by the World Energy Model, a large-scale simulation tool, developed at the IEA over a period of more than 20 years that is designed to replicate how energy markets function."

News & Hard to Categorize Content.

- [Budget credibility of subnational forecasts](#).

Budget credibility, or the ability of governments to accurately forecast macro-fiscal variables, is crucial for effective public finance management. Fiscal marksmanship analysis captures the extent of errors in the budgetary forecasting... Partitioning the sources of errors, we identified that the errors were more broadly random than due to systematic bias, except for a few crucial macro-fiscal variables where improving the forecasting techniques can provide better estimates.

- See also: [How accurate are \[US\] agencies' procurement forecasts?](#) and [Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods](#) (which finds random forests a hard to beat approach)
- [Bloomberg on the IMF's track record on forecasting \(archive link, without a paywall\)](#).

A Bloomberg analysis of more than 3,200 same-year country forecasts published each spring since 1999 found a wide variation in the direction and magnitude of errors. In 6.1 percent of cases, the IMF was within a 0.1 percentage-point margin of error. The rest of the time, its forecasts underestimated GDP growth in 56 percent of cases and overestimated it in 44 percent. The average forecast miss, regardless of direction, was 2.0 percentage points, but obscures a notable difference between the average 1.3 percentage-point error for advanced economies compared with 2.1 percentage points for more volatile and harder-to-model developing economies. Since the financial crisis, however, the IMF's forecast accuracy seems to have improved, as growth numbers have generally fallen.

Banking and sovereign debt panics hit Greece, Ireland, Portugal and Cyprus to varying degrees, threatening the integrity of the euro area and requiring emergency intervention from multinational authorities. During this period, the IMF wasn't merely forecasting what would happen to these countries but also setting the terms. It provided billions in bailout loans in exchange for implementation of strict austerity measures and other policies, often bitterly opposed by the countries' citizens and politicians.

- I keep seeing evidence that Trump will lose reelection, but I don't know how seriously to take it, because I don't know how filtered it is.
- For example, the [The Economist's model](#) forecasts 91% that Biden will win the upcoming USA elections. Should I update somewhat towards Biden winning after seeing it? What if I suspect that it's the most extreme model, and that it has come to my attention because of that fact? What if I suspect that it's the most extreme model which will predict a democratic win? What if there was another equally reputable model which predicts 91% for Trump, but which I never got to see because of information filter dynamics?
- The [the Primary Model](#) confirmed my suspicions of filter dynamics. It "does not use presidential approval or the state of the economy as predictors. Instead it relies on the performance of the presidential nominees in primaries", and on how many terms the party has controlled the White House. The model has been developed by an [otherwise unremarkable](#) professor of political science at New York's Stony Brook University, and has done well in previous election cycles. It assigns 91% to Trump winning reelection.
- [Forecasting at Uber: An Introduction](#). Uber forecasts demand so that they know amongst other things, when and where to direct their vehicles. Because of the

challenges to testing and comparing forecasting frameworks at scale, they developed their own software for this.

- [Forecasting Sales In These Uncertain Times.](#)

[...] a company selling to lower-income consumers might use the monthly employment report for the U.S. to see how people with just a high school education are doing finding jobs. A business selling luxury goods might monitor the stock market.

- [Unilever Chief Supply Officer on forecasting](#): "Agility does trump forecasting. At the end of the day, every dollar we spent on agility has probably got a 10x return on every dollar spent on forecasting or scenario planning."

An emphasis on agility over forecasting meant shortening planning cycles — the company reduced its planning horizon from 13 weeks to four. The weekly planning meeting became a daily meeting. Existing demand baselines and even artificial intelligence programs no longer applied as consumer spending and production capacity strayed farther from historical trends.

- [An updated introduction to prediction markets](#), yet one which contains some nuggets I didn't know about.

This bias toward favorable outcomes... appears for a wide variety of negative events, including diseases such as cancer, natural disasters such as earthquakes and a host of other events ranging from unwanted pregnancies and radon contamination to the end of a romantic relationship. It also emerges, albeit less strongly, for positive events, such as graduating from college, getting married and having favorable medical outcomes.

Nancy Reagan hired an astrologer, Joan Quigley, to screen Ronald Reagan's schedule of public appearances according to his horoscope, allegedly in an effort to avoid assassination attempts.

Google, Yahoo!, Hewlett-Packard, Eli Lilly, Intel, Microsoft, and France Telecom have all used internal prediction markets to ask their employees about the likely success of new drugs, new products, future sales.

Although prediction markets can work well, they don't always. IEM, PredictIt, and the other online markets were wrong about Brexit, and they were wrong about Trump's win in 2016. As the Harvard Law Review points out, they were also wrong about finding weapons of mass destruction in Iraq in 2003, and the nomination of John Roberts to the U.S. Supreme Court in 2005. There are also plenty of examples of small groups reinforcing each other's moderate views to reach an extreme position, otherwise known as groupthink, a theory devised by Yale psychologist Irving Janis and used to explain the Bay of Pigs invasion.

although thoughtful traders should ultimately drive the price, that doesn't always happen. The [prediction] markets are also no less prone to being caught in an information bubble than British investors in the South Sea Company in 1720 or speculators during the tulip mania of the Dutch Republic in 1637.

- [Food Supply Forecasting Company gets \\$12 million in Series A funding](#)

Long Content.

- [Michael Story](#), "Jotting down things I learned from being a superforecaster."

Small teams of smart, focused and rational generalists can absolutely smash big well-resourced institutions at knowledge production, for the same reasons startups can beat big rich incumbent businesses

There's a *lot* more to making predictive accuracy work in practice than winning a forecasting tournament. Competitions are about daily fractional updating, long lead times and exhaustive pre-forecast research on questions especially chosen for competitive suitability

Real life forecasting often requires fast turnaround times, fuzzy questions, and difficult-to-define answers with unclear resolution criteria. In a competition, a question with ambiguous resolution is thrown out, but in a crisis it might be the most important work you do

- Lukas Gloor on [takeaways from Covid forecasting on Metaculus](#)
- [Ambiguity aversion](#). "Better the devil you know than the devil you don't."

An ambiguity-averse individual would rather choose an alternative where the probability distribution of the outcomes is known over one where the probabilities are unknown. This behavior was first introduced through the [Ellsberg paradox](#) (people prefer to bet on the outcome of an urn with 50 red and 50 blue balls rather than to bet on one with 100 total balls but for which the number of blue or red balls is unknown).

- Gregory Lewis: [Use uncertainty instead of imprecision.](#)

If your best guess for X is 0.37, but you're very uncertain, you still shouldn't replace it with an imprecise approximation (e.g. "roughly 0.4", "fairly unlikely"), as this removes information. It is better to offer your precise estimate, alongside some estimate of its resilience, either subjectively ("0.37, but if I thought about it for an hour I'd expect to go up or down by a factor of 2"), or objectively ("0.37, but I think the standard error for my guess to be ~0.1").

- [Expert Forecasting with and without Uncertainty Quantification and Weighting: What Do the Data Say?](#): "it's better to combine expert uncertainties (e.g. 90% confidence intervals) than to combine their point forecasts, and it's better still to combine expert uncertainties based on their past performance."
 - See also a [1969 paper](#) by future Nobel Prize winner Clive Granger: "Two separate sets of forecasts of airline passenger data have been combined to form a composite set of forecasts. The main conclusion is that the composite set of forecasts can yield lower mean-square error than either of the original forecasts. Past errors of each of the original forecasts are used to determine the weights to attach to these two original forecasts in forming the combined forecasts, and different methods of deriving these weights are examined".
- [How to build your own weather forecasting model](#). Sailors realize that weather forecasting are often corrupted by different considerations (e.g., a reported 50% of rain doesn't happen 50% of the time), and search for better sources. One such source is the original, raw data used to generate weather forecasts: GRIB files (Gridded Information in Binary), which lack interpretation. But these have their own pitfalls, which sailors must learn to take into account. For example, GRIB files only take into account wind speed, not tidal acceleration, which can cause a significant increase in apparent wind.

'Forecasts are inherently political,' says Dashew. 'They are the result of people perhaps getting it wrong at some point so some pressures to interpret them in a different or more conservative way very often. These pressures change all the time so they are often subject to outside factors.'

Singleton says he understands how pressures on forecasters can lead to this opinion being formed: 'In my days at the Met Office when the Shipping Forecast used to work under me, they always said they try to tell it like it is and they do not try to make it sound worse.'

- [Forecasting the dividends of conflict prevention from 2020 - 2030](#). Study quantifies the dynamics of conflict, building a transition matrix between different states (peace, high risk, negative peace, war, and recovery) and validating it using historical dataset; they find (concurring with the previous literature), that countries have a tendency to fall into cycles of conflict. They conclude that changing this transition matrix would have a very high impact. Warning: extensive quoting follows.

Notwithstanding the mandate of the United Nations to promote peace and security, many member states are still sceptical about the dividends of conflict prevention. Their diplomats argue that it is hard to justify investments without being able to show its tangible returns to decision-makers and taxpayers. As a result, support for conflict prevention is halting and uneven, and governments and international agencies end up spending enormous sums in stability and peace support operations after-the-fact.

This study considers the trajectories of armed conflict in a 'business-as-usual' scenario between 2020-2030. Specifically, it draws on a comprehensive historical dataset to determine the number of countries that might experience rising levels of collective violence, outright armed conflict, and their associated economic costs. It then simulates alternative outcomes if conflict prevention measures were 25%, 50%, and 75% more effective. As with all projections, the quality of the projections relies on the integrity of the underlying data. The study reviews several limitations of the analysis, and underlines the importance of a cautious interpretation of the findings.

If current trends persist and no additional conflict prevention action is taken above the current baseline, then it is expected that there will be three more countries at war and nine more countries at high risk of war by 2030 as compared to 2020. This translates into roughly 677,250 conflict-related fatalities (civilian and battle-deaths) between the present and 2030. By contrast, under our most pessimistic scenario, a 25% increase in effectiveness of conflict prevention would result in 10 more countries at peace by 2030, 109,000 fewer fatalities over the next decade and savings of over \$3.1 trillion. A 50% improvement would result in 17 additional countries at peace by 2030, 205,000 fewer deaths by 2030, and some \$6.6 trillion in savings.

Meanwhile, under our most optimistic scenario, a 75% improvement in prevention would result in 23 more countries at peace by 2030, resulting in 291,000 lives saved over the next decade and \$9.8 trillion in savings. These scenarios are approximations, yet demonstrate concrete and defensible estimates of both the benefits (saved lives, displacement avoided, declining peacekeeping deployments) and cost-effectiveness of prevention (recovery aid, peacekeeping expenditures). Wars are costly and the avoidance of "conflict traps" could save the economy trillions of dollars by 2030 under the most optimistic scenarios. The bottom line is that comparatively modest investments in prevention can yield lasting effects by avoiding compounding costs of lost life, peacekeeping, and aid used for humanitarian response and rebuilding rather than development. The longer conflict prevention is delayed, the more expensive responses to conflict become.

In order to estimate the dividends of conflict prevention we analyze violence dynamics in over 190 countries over the period 1994 to 2017, a time period for which most data was available for most countries. Drawing on 12 risk variables, the model examines the likelihood that a war will occur in a country in the following year and we estimate (through linear, fixed effects regressions) the average cost of war (and other 'states', described below) on 8 dependent variables, including loss of life, displacement, peacekeeping deployments and expenditures, oversea aid and economic growth. The estimates confirm that, by far, the most costly state for a country to be in is war, and the probability of a country succumbing to war in the next year is based on its current state and the frequency of other countries with similar states having entered war in the past.

At the core of the model (and results) is the reality that countries tend to get stuck in so-called violence and conflict traps. A well-established finding in the conflict studies field is that once a country experiences an armed conflict, it is very likely to relapse into conflict or violence within a few years. Furthermore, countries likely to experience war share some common warning signs, which we refer to as "flags" (up to 12 flags can be raised to signal risk). Not all countries that enter armed conflict raise the same warning flags, but the warning flags are nevertheless a good indication that a country is at high risk. These effects create vicious cycles that result in high risk, war and frequent relapse into conflict. Multiple forms of prevention are necessary to break these cycles. The model captures the vicious cycle of conflict traps, through introducing five states and a transition matrix based on historical data (see Table 1). First, we assume that a country is in one of five 'states' in any given year. These 'states' are at "Peace", "High Risk", "Negative Peace", "War" and "Recovery" (each state is described further below). Drawing on historical data, the model assesses the probability of a country transitioning to another state in a given year (a transition matrix).

For example, if a state was at High Risk in the last year, it has a 19.3% chance of transitioning to Peace, a 71.4% chance of staying High Risk, a 7.6% chance of entering Negative Peace and a 1.7% chance of entering War the following year.

By contrast, high risk states are designated by the raising of up to 12 flags. These include: 1) high scores by Amnesty International's annual human rights reports (source: Political Terror Scale), 2) the US State Department annual reports (source: Political Terror Scale), 3) civilian fatalities as a percentage of population (source: ACLED), 4) political events per year (source: ACLED) 5) events attributed to the proliferation of non-state actors (source: ACLED), 6) battle deaths (source: UCDP), 7) deaths by terrorism (source: GTD), 8) high levels of crime (source: UNODC), 9) high levels of prison population (source: UNODC), 10) economic growth shocks (source: World Bank), 11) doubling of displacement in a year (source: IDMC), and 12) doubling of refugees in a year (source: UNHCR). Countries with two or more flags fall into the "high risk" category. Using these flags, a majority of countries have been at high risk for one or more years from 1994 to 2017, so it is easier to give examples of countries that have not been at high risk.

Negative peace states are defined by combined scores from Amnesty International and the US State Department. Countries in negative peace are more than five times as likely to enter high risk in the following year than peace (26.8% vs. 4.1%).

A country that is at war is one that falls into a higher threshold of collective violence, relative to the size of the population. Specifically, it is designated as such if one or more of the following conditions are met: above 0.04 battle deaths or .04 civilian fatalities per 100,000 according to UCDP and ACLED, respectively, or coding of genocide by the Political Instability Task Force Worldwide Atrocities Dataset.

Countries experiencing five or more years of war between 1994 and 2017 included Afghanistan, Somalia, Sudan, Iraq, Burundi, Central African Republic, Sri Lanka, DR Congo, Uganda, Chad, Colombia, Israel, Lebanon, Liberia, Yemen, Algeria, Angola, Sierra Leone, South Sudan, Eritrea and Libya.

Lastly, recovery is a period of stability that follows from war. A country is only determined to be recovering if it is not at war and was recently in a war. Any country that exits in the war state is immediately coded as being in recovery for the following five years, unless it relapses into war. The duration of the recovery period (five years) is informed by the work of Paul Collier et al, but is robust also to sensitivity tests around varying recovery lengths.

The model does not allow for countries to be high risk and in recovery in the same year, but there is ample evidence that countries that are leaving a war state are at a substantially higher risk of experiencing war recurrence, contributing to the conflict trap described earlier. Countries are twice as likely to enter high risk or negative peace coming out of recovery as they are to enter peace, and 10.2% of countries in recovery relapse into war every year. When a country has passed the five year threshold without reverting to war, it can move back to states of peace, negative peace or high risk.

The transition matrix underlines the very real risk of countries falling into a 'conflict trap'. Specifically, a country that is in a state of war has a very high likelihood of staying in this condition in the next year (72.6%) and just a 27.4% chance of transitioning to recovery. Once in recovery, a country has a 10.2% chance of relapse every year, suggesting only a 58% chance ($1-10.2\% \times 5$) that a country will not relapse over five years.

As Collier and others have observed, countries are often caught in prolonged and vicious cycles of war and recovery (conflict traps), often unable to escape into a new, more peaceful (or less war-like) state

- War is expensive. So is being at high risk of war.

Of course, the loss of life, displacement, and accumulated misery associated with war should be reason enough to invest in prevention, but there are also massive economic benefits from successful prevention. Foremost, the countries at war avoid the costly years in conflict, with growth rates 4.8% lower than countries at peace. They also avoid years of recovery and the risk of relapse into conflict. Where prevention works, conflict-driven humanitarian needs are reduced, and the international community avoids peacekeeping deployments and additional aid burdens, which are sizable.

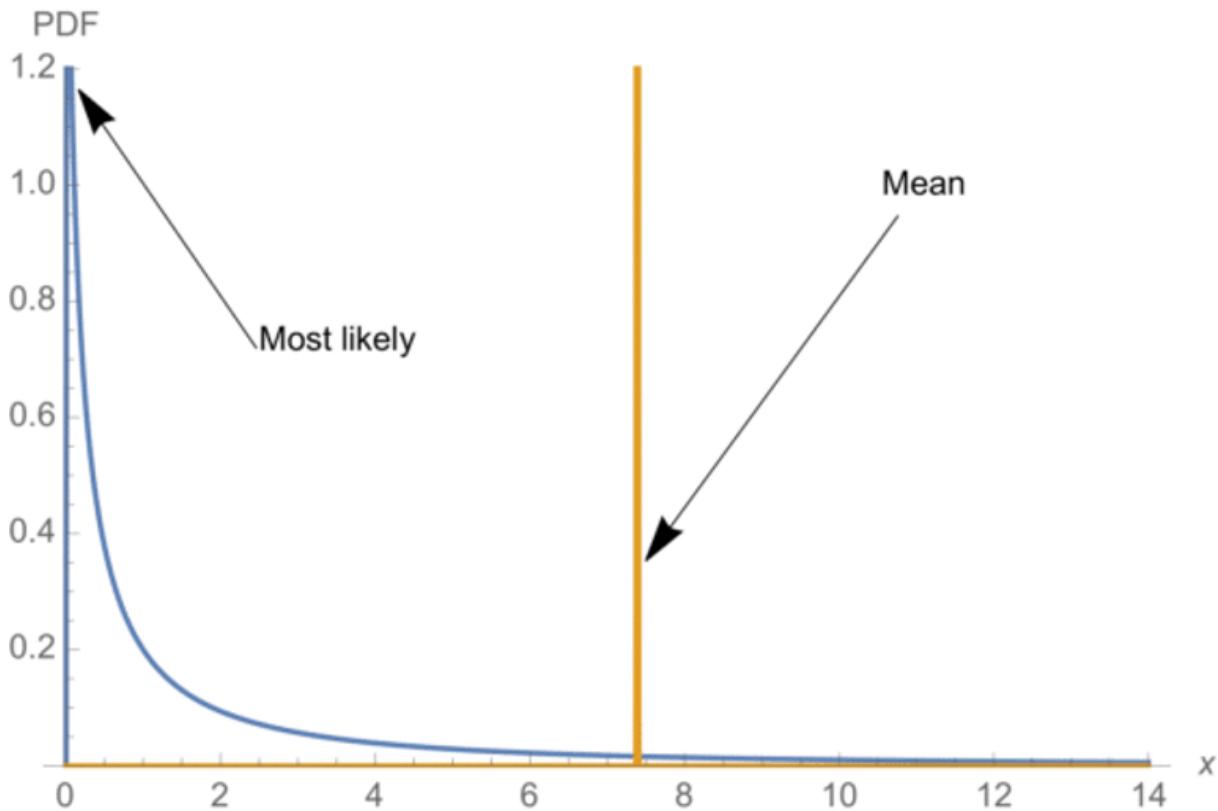
Conclusion: The world can be significantly better off by addressing the high risk of destructive violence and war with focused efforts at prevention in countries at high risk and those in negative peace. This group of countries has historically been at risk of higher conflict due to violence against civilians, proliferation of armed groups, abuses of human rights, forced displacement, high homicide, and incidence of error. None of this is surprising. Policymakers know that war is bad for humans and other living things. What is staggering is the annual costs of war that we will continue to pay in 2030 through inaction today – conceivably trillions of dollars of economic growth, and the associated costs of this for human security and development, are being swept off the table by the decisions made today to ignore prevention.

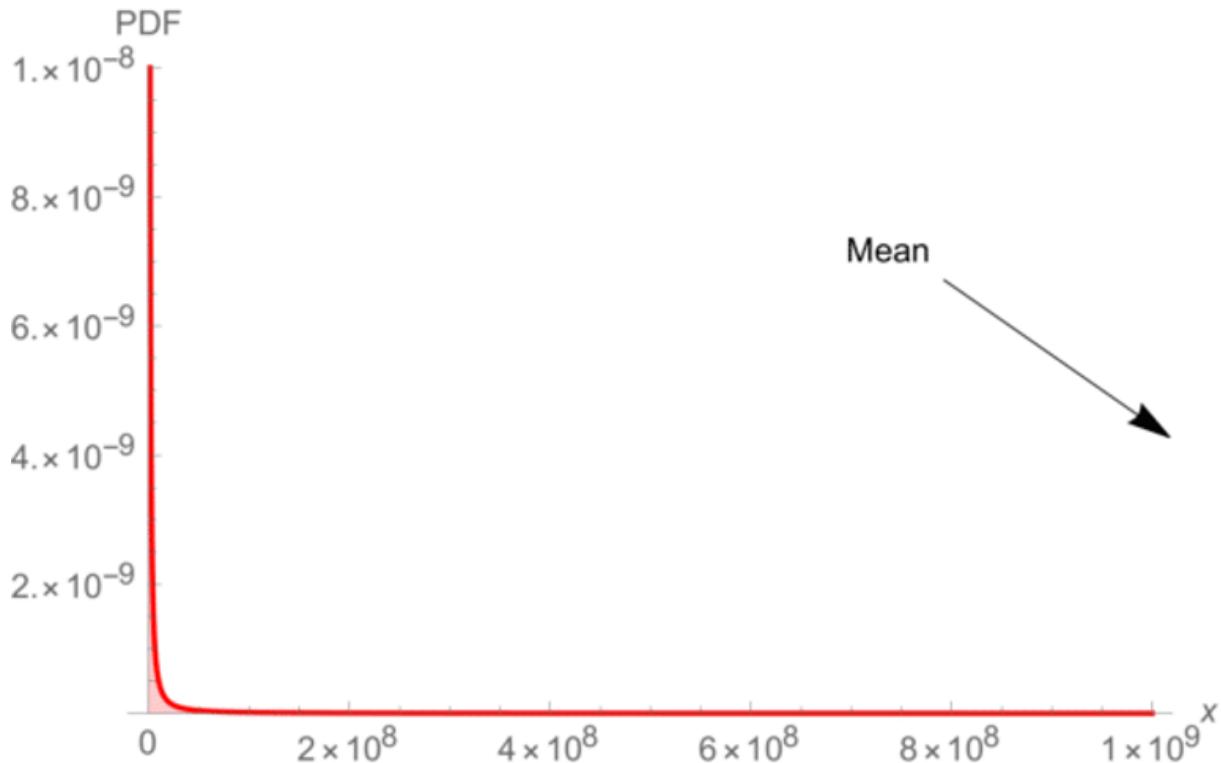
- [COVID-19: Ioannidis vs. Taleb](#)

On the one hand, Nassim Taleb has clearly expressed that measures to stop the spread of the pandemic must be taken as soon as possible: instead of looking at data, it is the nature of a pandemic with a possibility of devastating human impact that should drive our decisions.

On the other hand, John Ioannidis acknowledges the difficulty in having good data and of producing accurate forecasts, while believing that eventually any information that can be extracted from such data and forecasts should still be useful, e.g. to having targeted lockdowns (in space, time, and considering the varying risk for different segments of the population).

- [Taleb](#): *On single point forecasts for fat tailed variables.* Leitmotiv: Pandemics are fat-tailed.





We do not need more evidence under fat tailed distributions — it is there in the properties themselves (properties for which we have ample evidence) and these clearly represent risk that must be killed in the egg (when it is still cheap to do so). Secondly, unreliable data — or any source of uncertainty — should make us follow the most paranoid route. [...] more uncertainty in a system makes precautionary decisions very easy to make (if I am uncertain about the skills of the pilot, I get off the plane).

Random variables in the power law class with tail exponent $\alpha \leq 1$ are, simply, not forecastable. They do not obey the [Law of Large Numbers]. But we can still understand their properties.

As a matter of fact, owing to preasymptotic properties, a heuristic is to consider variables with up to $\alpha \leq 5/2$ as not forecastable — the mean will be too unstable and requires way too much data for it to be possible to do so in reasonable time. It takes 10^{14} observations for a “Pareto 80/20” (the most commonly referred to probability distribution, that is with $\alpha \approx 1.13$) for the average thus obtained to emulate the significance of a Gaussian with only 30 observations.

- [Ioannidis](#): *Forecasting for COVID-19 has failed*. Leitmotiv: "Investment should be made in the collection, cleaning and curation of data".

Predictions for hospital and ICU bed requirements were also entirely misinforming. Public leaders trusted models (sometimes even black boxes without disclosed methodology) inferring massively overwhelmed health care capacity (Table 1) [3]. However, eventually very few hospitals were stressed, for a couple of weeks. Most hospitals maintained largely empty wards, waiting for tsunamis that never came. The general population was locked and placed in horror-alert to save the health system from collapsing. Tragically, many health systems faced major adverse consequences, not by COVID-19 cases overload, but for very different reasons. Patients with heart

attacks avoided visiting hospitals for care [4], important treatments (e.g. for cancer) were unjustifiably delayed [5], mental health suffered [6]. With damaged operations, many hospitals started losing personnel, reducing capacity to face future crises (e.g. a second wave). With massive new unemployment, more people may lose health insurance. The prospects of starvation and of lack of control for other infectious diseases (like tuberculosis, malaria, and childhood communicable diseases for which vaccination is hindered by the COVID-19 measures) are dire...

The core evidence to support “flatten-the-curve” efforts was based on observational data from the 1918 Spanish flu pandemic on 43 US cities. These data are >100-years old, of questionable quality, unadjusted for confounders, based on ecological reasoning, and pertaining to an entirely different (influenza) pathogen that had ~100-fold higher infection fatality rate than SARS-CoV-2. Even thus, the impact on reduction on total deaths was of borderline significance and very small (10-20% relative risk reduction); conversely many models have assumed 25-fold reduction in deaths (e.g. from 510,000 deaths to 20,000 deaths in the Imperial College model) with adopted measures

Despite these obvious failures, epidemic forecasting continued to thrive, perhaps because vastly erroneous predictions typically lacked serious consequences. Actually, erroneous predictions may have been even useful. A wrong, doomsday prediction may incentivize people towards better personal hygiene. Problems starts when public leaders take (wrong) predictions too seriously, considering them crystal balls without understanding their uncertainty and the assumptions made. Slaughtering millions of animals in 2001 aggravated a few animal business stakeholders, most citizens were not directly affected. However, with COVID-19, espoused wrong predictions can devastate billions of people in terms of the economy, health, and societal turmoil at-large.

Cirillo and Taleb thoughtfully argue [14] that when it comes to contagious risk, we should take doomsday predictions seriously: major epidemics follow a fat-tail pattern and extreme value theory becomes relevant. Examining 72 major epidemics recorded through history, they demonstrate a fat-tailed mortality impact. However, they analyze only the 72 most noticed outbreaks, a sample with astounding selection bias. The most famous outbreaks in human history are preferentially selected from the extreme tail of the distribution of all outbreaks. Tens of millions of outbreaks with a couple deaths must have happened throughout time. Probably hundreds of thousands might have claimed dozens of fatalities. Thousands of outbreaks might have exceeded 1,000 fatalities. Most eluded the historical record. The four garden variety coronaviruses may be causing such outbreaks every year [15,16]. One of them, OC43 seems to have been introduced in humans as recently as 1890, probably causing a “bad influenza year” with over a million deaths [17]. Based on what we know now, SARS-CoV-2 may be closer to OC43 than SARS-CoV-1. This does not mean it is not serious: its initial human introduction can be highly lethal, unless we protect those at risk.

- The (British) Royal Economic Society presents a panel on [What is a scenario, projection and a forecast - how good or useful are they particularly now?](#). The start seems promising: "My professional engagement with economic and fiscal forecasting was first as a consumer, and then a producer. I spent a decade happily mocking other people's efforts, as a journalist, since when I've spent two decades helping colleagues to construct forecasts and to try to explain them to the public." The first speaker, which corresponds to the first ten minutes, is worth listening to; the rest varies in quality.

You have to construct the forecast and explain it in a way that's fit for that purpose

- I liked the following taxonomy of what distinct targets the agency the first speaker works for is aiming to hit with their forecasts:

1. as an input into the policy-making process,
2. as a transparent assessment of public finances
3. as a prediction of whether the government will meet whatever fiscal rules it has set itself,
4. as a baseline against which to judge the significance of further news,
5. as a challenge to other agencies "to keep the bastards honest".

- The limitations were interesting as well:

1. they require us to produce a forecast that's conditioned on current government policy even if we and everyone else expect that policy to change that of course makes it hard to benchmark our performance against counterparts who are producing unconditional forecasts.
2. The forecasts have to be explainable; a black box model might be more accurate but be less useful.
3. they require detailed discussion of the individual forecast lines and clear diagnostics to explain changes from one forecast to the next precisely to reassure people that those changes aren't politically motivated or tainted - the forecast is as much about delivering transparency and accountability as about demonstrating predictive prowess
4. the forecast numbers really have to be accompanied by a comprehensible narrative of what is going on in the economy and the public finances and what impact policy will have - Parliament and the public needs to be able to engage with the forecast we couldn't justify our predictions simply with an appeal to a statistical black box and the Chancellor certainly couldn't justify significant policy positions that way.

"horses for courses, the way you do the forecast, the way you present it depends on what you're trying to achieve with it"

"People use scenario forecasting in a very informal manner. which I think that could be problematic because it's very difficult to basically find out what are the assumptions and whether those assumptions and the models and the laws can be validated"

Linear models are state independent, but it's not the same to receive a shock where the economy is in upswing as when the economy is during a recession.

- Some situations are too complicated to forecast, so one conditions on some variables being known, or following a given path, and then studies the rest, calling the output a "scenario."

One week delay in intervention by the government makes a big difference to the height of the [covid-19] curve.

I don't think it's easy to follow the old way of doing things. I'm sorry, I have to be honest with you. I spent 4 months just thinking about this problem and you need to

integrate a model of the social behavior and how you deal with the risk to health and to economy in these models. But unfortunately, by the time we do that it won't be relevant.

It amuses me to look at weather forecasts because economists don't have that kind of technology, those kind of resources.

Note to the future: All links are added automatically to the Internet Archive. In case of link rot, go [here](#)

"horses for courses, the way you do the forecast, the way you present it depends on what you're trying to achieve with it"

Forecasting Newsletter: August 2020.

Highlights

538 releases [model](#) of the US elections; Trump predicted to win ~30% of the time.

[Study](#) offers instructive comparison of New York covid models, finds that for the IHME model, reported death counts fell inside the 95% prediction intervals only 53% of the time.

Bigest decentralized trial [to date](#), with 511 jurors asked to adjudicate a case coming from the Omen prediction market: "Will there be a day with at least 1000 reported corona deaths in the US in the first 14 days of July?."

Index

- Highlights
- Prediction Markets & Forecasting Platforms
- In The News
- Hard To Categorize
- Long Content

Sign up [here](#) or browse past newsletters [here](#).

Prediction Markets & Forecasting Platforms

On [PredictIt](#), presidential election prices are close to [even odds](#), with Biden at 55, and Trump at 48.

Good Judgement Inc. continues providing their [dashboard](#), and the difference between the probability assigned by superforecasters to a Biden win (~75%), and those offered by [betfair](#) (~55%) was enough to make it worth for me to place a small bet. At some point, Good Judgement Inc. and Cultivate Labs started a new platform on the domain [covidimpacts.com](#), but forecasts there seem weaker than on Good Judgement Open.

[Replication Markets](#) started their COVID-19 round, and created a page with [COVID-19 resources for forecasters](#).

Nothing much to say about [Metaculus](#) this month, but I appreciated their previously existing list of [prediction resources](#).

[Foretell](#) has a [blog](#), and hosted a forecasting forum which discussed

- metrizing the grand. That is, decomposing and operationalizing big picture questions into smaller ones, which can then be forecasted.
- operationalizing these big picture questions might also help identify disagreements, which might then either be about the indicators, proxies or subquestions chosen, or about the probabilities given to the subquestions.

- sometimes we can't measure what we care about, or we don't care about what we can measure.
- one might be interested in questions about the future 3 to 7 years from now, but questions which ask about events 3 to 15 months in the future (which forecasting tournaments can predict better) can still provide useful signposts.

Meanwhile, ethereum-based prediction markets such as Omen or Augur are experiencing difficulties because of the rise of decentralized finance (DeFi) and speculation and excitement about it. That speculation and excitement has increased the gas price (fees), such that making a casual prediction is for now too costly.

In The News

[Forecasting the future of philanthropy](#). The [American Lebanese Syrian Associated Charities](#), the largest healthcare related charity in the United States, whose mission is to fund the [St. Jude Children's Research Hospital](#). To do this, they employ aggressive fundraising tactics, which have undergone modifications throughout the current pandemic.

[Case 302: the Largest Decentralized Trial of All Time](#). Kleros is a decentralized dispute resolution platform. "In July, Kleros had its largest trial ever where 511 jurors were drawn in the General Court to adjudicate a case coming from the Omen prediction market: Will there be a day with at least 1000 reported Corona death in the US in the first 14 days of July?" [Link to the case](#)

[ExxonMobil Slashing Permian Rig Count, Forecasting Global Oil Glut Extending 'Well into 2021'](#). My own interpretation is that the gargantuan multinational's decision is an honest signal of an expected extended economic downturn.

Supply is expected to exceed demand for months, "and we anticipate it will be well into 2021 before the overhang is cleared and we returned to pre-pandemic levels," Senior Vice President Neil Chapman said Friday during a conference call.

"Simply put, the demand destruction in the second quarter was unprecedented in the history of modern oil markets. To put it in context, absolute demand fell to levels we haven't seen in nearly 20 years. We've never seen a decline with this magnitude and pace before, even relative to the historic periods of demand volatility following the global financial crisis and as far back as the 1970s oil and energy crisis."

Even so, ExxonMobil's Permian rig count is to be sharply lower than it was a year ago. The company had more than 50 rigs running across its Texas-New Mexico stronghold as of last fall. At the end of June it was down to 30, "and we expect to cut that number by at least half again by the end of this year," Chapman said.

[Google Cloud AI and Harvard Global Health Institute Collaborate on new COVID-19 forecasting model](#).

[Betting markets](#) put [UK-EU trade deal in 2020 at 66%](#) (now 44%).

[Experimental flood forecasting system didn't help](#) in Mumbai. The system was to provide a three day advance warning, but didn't.

FiveThirtyEight covers various facets of the USA elections: [Biden Is Polling Better Than Clinton At Her Peak](#), and releases [their model](#), along with some [comments about it](#)

In other news, this newsletter reached 200 subscribers last week.

Hard to Categorize

[Groundhog day](#) is a tradition in which American crowds pretend to believe that a small rat has oracular powers.

[Tips](#) for forecasting on PredictIt. These include betting against Trump voters who arrive at PredictIt from Breitbart.

Linch Zhang asks [What are some low-information priors that you find practically useful for thinking about the world?](#)

[AstraZeneca looking for a Forecasting Director](#) (US-based).

[Genetic Engineering Attribution Challenge](#).

NSF-funded tournament looking to compare human forecasters with a random forest ML model from Johns Hopkins in terms of forecasting the success probability of cancer drug trials. More info [here](#), and one can sign-up [here](#). I've heard rewards are generous, but they don't seem to be specified on the webpage. Kudos to Joshua Monrad.

Results of an [expert forecasting session](#) on covid, presented by expert forecaster Juan Cambeiro.

A playlist of [podcasts related to forecasting](#). Kudos to Michał Dubrawski.

Long Content

[A case study in model failure? COVID-19 daily deaths and ICU bed utilization predictions in New York state](#) and commentary: [Individual model forecasts can be misleading, but together they are useful](#).

In this issue, Chin et al. compare the accuracy of four high profile models that, early during the outbreak in the US, aimed to make quantitative predictions about deaths and Intensive Care Unit (ICU) bed utilization in New York. They find that all four models, though different in approach, failed not only to accurately predict the number of deaths and ICU utilization but also to describe uncertainty appropriately, particularly during the critical early phase of the epidemic. While overcoming these methodological challenges is key, Chin et al. also call for systemic advances including improving data quality, evaluating forecasts in real-time before policy use, and developing multi-model approaches.

But what the model comparison by Chin et al. highlights is an important principle that many in the research community have understood for some time: that no single model should be used by policy makers to respond to a rapidly changing, highly uncertain epidemic, regardless of the institution or modeling group from which it comes. Due to the multiple uncertainties described above, even models using the same underlying data often have results that diverge because they have

made different but reasonable assumptions about highly uncertain epidemiological parameters, and/or they use different methods

.. the rapid deployment of this approach requires pre-existing infrastructure and evaluation systems now and for improved response to future epidemics. Many models that are built to forecast on a scale useful for local decision making are complex, and can take considerable time to build and calibrate

a group with a history of successful influenza forecasting in the US (Los Alamos National Lab (4)) was able to produce early COVID-19 forecasts and had the best coverage of uncertainty in the Chin et al. analysis (80-100% of observations fell within the 95% prediction interval for most forecasts). In contrast, the new Institute for Health Metrics and Evaluation statistical approach had low reliability; after the latest analyzed revision only 53% of reported death counts fell with the 95% prediction intervals.

The original IHME model underestimates uncertainty and 45.7% of the predictions (over 1- to 14-step-ahead predictions) made over the period March 24 to March 31 are outside the 95% PIs. In the revised model, for forecasts from April 3 to May 3 the uncertainty bounds are enlarged, and most predictions (74.0%) are within the 95% PIs, which is not surprising given the PIs are in the order of 300 to 2000 daily deaths. Yet, even with this major revision, the claimed nominal coverage of 95% well exceeds the actual coverage. On May 4, the IHME model undergoes another major revision, and the uncertainty is again dramatically reduced with the result that 47.4% of the actual daily deaths fall outside the 95% PIs—well beyond the claimed 5% nominal value.

the LANL model was the only model that was found to approach the 95% nominal coverage, but unfortunately this model was unavailable at the time Governor Cuomo needed to make major policy decisions in late March 2020.

Models that are consistently poorly performing should carry less weight in shaping policy considerations. Models may be revised in the process, trying to improve performance. However, improvement of performance against retrospective data offers no guarantee for continued improvement in future predictions. Failed and recast models should not be given much weight in decision making until they have achieved a prospective track record that can instill some trust for their accuracy. Even then, real time evaluation should continue, since a model that performed well for a given period of time may fail to keep up under new circumstances.

[Do Prediction Markets Produce Well-Calibrated Probability Forecasts?](#)

Abstract: This article presents new theoretical and empirical evidence on the forecasting ability of prediction markets. We develop a model that predicts that the time until expiration of a prediction market should negatively affect the accuracy of prices as a forecasting tool in the direction of a ‘favourite/longshot bias’. That is, high-likelihood events are underpriced, and low-likelihood events are over-priced. We confirm this result using a large data set of prediction market transaction prices. Prediction markets are reasonably well calibrated when time to expiration is relatively short, but prices are significantly biased for events farther in the future. When time value of money is considered, the miscalibration can be exploited to earn excess returns only when the trader has a relatively low discount rate.

We confirm this prediction using a data set of actual prediction markets prices from 1,787 market representing a total of more than 500,000 transactions

Paul Christiano on [learning the Prior](#) and on [better priors as a safety problem](#).

A presentation of [radical probabilism](#); a theory of probability which relaxes some assumptions in classical Bayesian reasoning.

[Forecasting Thread: AI timelines](#), which asks for (quantitative) forecasts until human-machine parity. Some of the answers seem insane or suspicious, in that they have very narrow tails, sharp spikes, and don't really update on the fact that other people disagree with them.

Note to the future: All links are added automatically to the Internet Archive. In case of link rot, go [there](#) and input the dead link.

We hope that people will pressure each other into operationalizing their [big picture outlooks]. If we have no way of proving you wrong, we have no way of proving you right. We need falsifiable forecasts.

Source: Foretell Forecasting Forum. Inexact quote.

Forecasting Newsletter: September 2020.

Highlights

- Red Cross and Red Crescent societies have been trying out [forecast based financing](#), where funds are released before a potential disaster happens based on forecasts thereof.
- Andrew Gelman releases [Information, incentives, and goals in election forecasts](#); 538's 80% political predictions turn out to have happened [88% of the time](#).
- Nonprofit Ought organizes a [forecasting thread on existential risk](#), where participants display and discuss their probability distributions for existential risk.

Index

- Highlights
- Prediction Markets & Forecasting Platforms
- In The News
- Hard To Categorize
- Long Content

Sign up [here](#) or browse past newsletters [here](#).

Prediction Markets & Forecasting Platforms

Metaculus updated their [track record page](#). You can now look at accuracy across time, at the distribution of brier scores, and a calibration graph. They also have a new black swan question: [When will US metaculus users face an emigration crisis?](#).

Good Judgement Open has a [thread](#) in which forecasters share and discuss tips, tricks and experiences. An account is needed to browse it.

[Augur](#) modifications in response to higher ETH prices. Some unfiltered comments [on reddit](#)

An overview of [PlotX](#), a new decentralized prediction protocol/marketplace. PlotX focuses on non-subjective markets that can be programmatically determined, like the exchange rate between currencies or tokens.

A Replication Markets participant wrote [What's Wrong with Social Science and How to Fix It: Reflections After Reading 2578 Papers](#). See also: [An old long-form introduction to Replication Markets](#).

Georgetown's CSET is attempting to use forecasting to influence policy. A seminar discussing their approach [Using Crowd Forecasting to Inform Policy with Jason Matheny](#) is scheduled for the 19th of October. But their current forecasting tournament, foretell, isn't yet very well populated, and the aggregate isn't that good because participants don't update all that often, leading to sometimes clearly outdated aggregates. Perhaps

because of this relative lack of competition, my team is in 2nd place at the time of this writting (with myself at #6, Eli Lifland at #12 and Misha Yagudin at #21). You can join foretell [here](#).

There is a new contest on Hypermind, [The Long Fork Project](#), which aims to predict the impact of a Trump or a Biden victory in November, with \$20k in prize money. H/t to user [ChickCounterfly](#).

The University of Chicago's Effective Altruism group is hosting a forecasting tournament between all interested EA college groups starting October 12th, 2020. More details [here](#)

In the News

News media sensationalizes essentially random fluctuations on US election odds caused by big bettors entering prediction markets such as Betfair, where bets on the order of \$50k can visibly alter the market price. Simultaneously, polls/models and prediction market odds have diverged, because a substantial fraction of bettors lend credence to the thesis that polls will be biased as in the previous elections, even though polling firms seem to have improved their methods.

- [Trump overtakes Biden as favorite to win in November: Betfair Exchange](#)
- [US Election: Polls defy Trump's comeback narrative but will the market react?](#)
- [Betting Markets Swing Toward Trump, Forecasting Tightening Race](#)
- [Biden leads in the polls, but betters are taking a gamble on Trump](#)
- [UK Bookmaker Betfair Shortens Joe Biden 2020 Odds After Bettor Wagers \\$67K](#)
- [Avoid The Monster Trump Gamble - The Fundamental Numbers Haven't Changed](#)

Red Cross and Red Crescent societies have been trying out forecast based financing. The idea is to create forecasts and early warning indicators for some negative outcome, such as a flood, using weather forecasts, satellite imagery, climate models, etc, and then release funds automatically if the forecast reaches a given threshold, allowing the funds to be put to work before the disaster happens in a more automatic, fast and efficient manner. Goals and modus operandi might resonate with the Effective Altruism community: > "In the precious window of time between a forecast and a potential disaster, FbF releases resources to take early action. Ultimately, we hope this early action will be more **effective at reducing suffering**, compared to waiting until the disaster happens and then doing only disaster response. For example, in Bangladesh, people who received a forecast-based cash transfer were less malnourished during a flood in 2017." (bold not mine)

- Here is the "what can go wrong" section of their [slick yet difficult to navigate webpage](#), and an introductory [video](#).

[Prediction Markets' Time Has Come, but They Aren't Ready for It](#). Prediction markets could have been useful for predicting the spread of the pandemic (see: [coronainformationmarkets.com](#)), or for informing presidential election consequences (see: Hypermind above), but their relatively small size makes them less informative. Blockchain based prediction technologies, like Augur, Gnosis or Omen could have helped bypass US regulatory hurdles (which ban many kinds of gambling), but the recent increase in transaction fees means that "everything below a \$1,000 bet is basically economically unfeasible"

Floods in India and Bangladesh:

- [Time to develop a reliable flood forecasting model \(for Bangladesh\)](#)

This year, flood started somewhat earlier than usual. The Brahmaputra water crossed the danger level (DL) on June 28, subsided after a week, and then crossed the DL again on July 13 and continued for 26 days. It inundated over 30 percent of the country

- [Google's AI Flood Forecasting Initiative now expanded to all parts of India; Google bolsters its A.I.-enabled flood alerts for India and Bangladesh](#)

"One assumption that was presumed to be true in hydrology is that you cannot generalize across water basins," Nevo said. "Well, it's not true, as it turns out." He said Google's A.I.-based forecasting model has performed better on watersheds it has never encountered before in training than classical hydrologic models that were designed specifically for that river basin.

[The many tribes of 2020 election worriers: An ethnographic report by the Washington Post.](#)

Electricity time series demand and supply forecasting startup [raises \\$8 million](#). I keep seeing this kind of announcement; doing forecasting well in an underforecasted domain seems to be somewhat profitable right now, and it's not like there is an absence of domains to which forecasting can be applied. This might be a good idea for an earning-to-give startup.

[NSF and NASA partner to address space weather research and forecasting](#). Together, NSF and NASA are investing over \$17 million into six, three-year awards, each of which contributes to key research that can expand the nation's space weather prediction capabilities.

In its monthly report, OPEC said it expects the pandemic to reduce demand by 9.5 million barrels a day, forecasting a fall in demand of 9.5% from last year, [reports the Wall Street Journal](#)

Some [criticism](#) of Gnosis, a decentralized prediction markets startup, by early investors who want to cash out. [Here](#) is a blog post by said early investors; they claim that "Gnosis took out what was in effect a 3+ year interest-free loan from token holders and failed to deliver the products laid out in its fundraising whitepaper, quintupled the size of its balance sheet due simply to positive price fluctuations in ETH, and then launched products that accrue value only to Gnosis management."

[What a study of video games can tell us about being better decision makers](#) (\$), a frustratingly well-paywalled, yet exhaustive, complete and informative overview of the IARPA's FOCUS tournament:

To study what makes someone good at thinking about counterfactuals, the intelligence community decided to study the ability to forecast the outcomes of simulations. A simulation is a computer program that can be run again and again, under different conditions: essentially, rerunning history. In a simulated world, the researchers could know the effect a particular decision or intervention would have. They would show teams of analysts the outcome of one run of the simulation and then ask them to predict what would have happened if some key variable had been changed.

Negative Examples

[Why Donald Trump Isn't A Real Candidate, In One Chart](#), wrote 538 in 2015.

For this reason alone, Trump has a better chance of cameoing in another "Home Alone" movie with Macaulay Culkin — or playing in the NBA Finals — than winning the Republican nomination.

[Travel CFOs Hesitant on Forecasts as Pandemic Fogs Outlook](#), reports the Wall Street Journal.

"We're basically prevented from saying the word 'forecast' right now because whatever we forecast...it's wrong," said Shannon Okinaka, chief financial officer at Hawaiian Airlines. "So we've started to use the word 'planning scenarios' or 'planning assumptions.'"

Long Content

Andrew Gelman et al. release [Information, incentives, and goals in election forecasts](#).

- Neither The Economist's model nor 538's are fully Bayesian. In particular, they are not martingales, that is, their current probability is not the expected value of their future probability.

campaign polls are more stable than every before, and even the relatively small swings that do appear can largely be attributed to differential nonresponse

Regarding predictions for 2020, the creator of the FiveThirtyEight forecast writes, "we think it's appropriate to make fairly conservative choices especially when it comes to the tails of your distributions. Historically this has led 538 to well-calibrated forecasts (our 20% really mean 20%)" (Silver, 2020b). But conservative prediction corresponds can produce a too-wide interval, one that plays it safe by including extra uncertainty. In other words, conservative forecasts should lead to underconfidence: intervals whose coverage is greater than advertised. And, indeed, according to the calibration plot shown by Boice and Wezerek (2019) of FiveThirtyEight's political forecasts, in this domain 20% for them really means 14%, and 80% really means 88%.

[The Literary Digest Poll of 1936](#). A poll so bad that it destroyed the magazine.

- Compare the Literary Digest and Gallup polls of 1936 with The New York Times's [model of 2016](#) and [538's 2016 forecast](#), respectively.

In retrospect, the polling techniques employed by the magazine were to blame. Although it had polled ten million individuals (of whom 2.27 million responded, an astronomical total for any opinion poll),^[5] it had surveyed its own readers first, a group with disposable incomes well above the national average of the time (shown in part by their ability to afford a magazine subscription during the depths of the Great Depression), and those two other readily available lists, those of registered automobile owners and that of

telephone users, both of which were also wealthier than the average American at the time.

Research published in 1972 and 1988 concluded that as expected this sampling bias was a factor, but non-response bias was the primary source of the error - that is, people who disliked Roosevelt had strong feelings and were more willing to take the time to mail back a response.

George Gallup's American Institute of Public Opinion achieved national recognition by correctly predicting the result of the 1936 election, while Gallup also correctly predicted the (quite different) results of the Literary Digest poll to within 1.1%, using a much smaller sample size of just 50,000. [5] Gallup's final poll before the election also predicted Roosevelt would receive 56% of the popular vote: the official tally gave Roosevelt 60.8%.

This debacle led to a considerable refinement of public opinion polling techniques, and later came to be regarded as ushering in the era of modern scientific public opinion research.

[Feynman in 1985](#), answering questions about whether machines will ever be more intelligent than humans.

[Why Most Published Research Findings Are False](#), back from 2005. The abstract reads:

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

[Reference class forecasting](#). Reference class forecasting or comparison class forecasting is a method of predicting the future by looking at similar past situations and their outcomes. The theories behind reference class forecasting were developed by Daniel Kahneman and Amos Tversky. The theoretical work helped Kahneman win the Nobel Prize in Economics. Reference class forecasting is so named as it predicts the outcome of a planned action based on actual outcomes in a reference class of similar actions to that being forecast.

[Reference class problem](#)

In statistics, the reference class problem is the problem of deciding what class to use when calculating the probability applicable to a particular case. For example, to estimate the probability of an aircraft crashing, we could refer to the frequency of crashes among various different sets of aircraft: all aircraft, this make of aircraft, aircraft flown by this company in the last ten years, etc. In this example, the

aircraft for which we wish to calculate the probability of a crash is a member of many different classes, in which the frequency of crashes differs. It is not obvious which class we should refer to for this aircraft. In general, any case is a member of very many classes among which the frequency of the attribute of interest differs. The reference class problem discusses which class is the most appropriate to use.

- See also some thoughts on this [here](#)

[The Base Rate Book](#) by Credit Suisse.

This book is the first comprehensive repository for base rates of corporate results. It examines sales growth, gross profitability, operating leverage, operating profit margin, earnings growth, and cash flow return on investment. It also examines stocks that have declined or risen sharply and their subsequent price performance. We show how to thoughtfully combine the inside and outside views. The analysis provides insight into the rate of regression toward the mean and the mean to which results regress.

Hard To Categorize

[Improving decisions with market information: an experiment on corporate prediction markets \(sci-hub; archive link\)](#)

We conduct a lab experiment to investigate an important corporate prediction market setting: A manager needs information about the state of a project, which workers have, in order to make a state-dependent decision. Workers can potentially reveal this information by trading in a corporate prediction market. We test two different market designs to determine which provides more information to the manager and leads to better decisions. We also investigate the effect of top-down advice from the market designer to participants on how the prediction market is intended to function. Our results show that the theoretically superior market design performs worse in the lab—in terms of manager decisions—without top-down advice. With advice, manager decisions improve and both market designs perform similarly well, although the theoretically superior market design features less mispricing. We provide a behavioral explanation for the failure of the theoretical predictions and discuss implications for corporate prediction markets in the field.

The nonprofit Ought organized a [forecasting thread on existential risk](#), where participants display and discuss their probability distributions for existential risk, and outline some [reflections on a previous forecasting thread on AI timelines](#).

A [draft report on AI timelines, summarized in the comments](#)

Gregory Lewis has a series of posts related to forecasting and uncertainty:

- [Use resilience, instead of imprecision, to communicate uncertainty](#)
- [Challenges in evaluating forecaster performance](#)
- [Take care with notation for uncertain quantities](#)

[Estimation of probabilities to get tenure track in academia: baseline and publications during the PhD.](#)

[How to think about an uncertain future: lessons from other sectors & mistakes of longtermist EAs](#). The central thesis is:

Expected value calculations, the favoured approach for EA decision making, are all well and good for comparing evidence backed global health charities, but they are often the wrong tool for dealing with situations of high uncertainty, the domain of EA longtermism.

Discussion by a PredictIt bettor on [how he made money by following Nate Silver's predictions](#), from r/TheMotte.

Also on r/TheMotte, on [the promises and deficiencies of prediction markets](#):

Prediction markets will never be able to predict the unpredictable. Their promise is to be better than all of the available alternatives, by incorporating all available information sources, weighted by experts who are motivated by financial returns.

So, you'll never have a perfect prediction of who will win the presidential election, but a good prediction market could provide the best possible guess of who will win the presidential election.

To reach that potential, you'd need to clear away the red tape. It would need to be legal to make bets on the market, fees for making transaction need to be low, participants would need faith in the bet adjudication process, and there can't be limits to the amount you can bet. Signs that you'd succeeded would include sophisticated investors making large bets with a narrow bid/ask spread.

Unfortunately prediction markets are nowhere close to that ideal today; they're at most "barely legal," bet sizes are limited, transaction fees are high, getting money in or out is clumsy and sketchy, trading volumes are pretty low, and you don't see any hedge funds with "prediction market" desks or strategies. As a result, I put very little stock in political prediction markets today. At best they're populated by dumb money, and at worst they're actively manipulated by campaigns or partisans who are not motivated by direct financial returns.

[Nate Silver](#) on a small twitter thread on prediction markets: "Most of what makes political prediction markets dumb is that people assume they have expertise about election forecasting because they a) follow politics and b) understand "data" and "markets". Without more specific domain knowledge, though, that combo is a recipe for stupidity."

- Interestingly, I've recently found out that 538's political predictions are probably [underconfident](#), i.e., an 80% happens 88% of the time.

[Deloitte](#) forecasts US holiday season retail sales (but doesn't provide confidence intervals.)

[Solar forecast](#). Sun to leave the quietest part of its cycle, but still remain relatively quiet and not produce world-ending coronal mass ejections, the New York Times reports.

The Foresight Institute organizes weekly talks; here is one with Samo Burja on [long-lived institutions](#).

[Some examples of failed technology predictions](#).

Last, but not least, Ozzie Gooen on [Multivariate estimation & the Squiggly language](#):



Distributions

Functions that return distributions

Note to the future: All links are added automatically to the Internet Archive. In case of link rot, go [there](#) and input the dead link.

[Littlewood's law](#) states that a person can expect to experience events with odds of one in a million (defined by the law as a "miracle") at the rate of about one per month."

Forecasting Newsletter: October 2020.

Highlights

- Facebook's Forecast now out of [out of beta](#).
- British Minister and experts give [probabilistic predictions](#) of the chance of a Brexit deal.
- CSET/Foretell publishes an [issue brief](#) on their approach to using forecasters to inform big picture policy questions.

Index

- Highlights
- Prediction Markets & Forecasting Platforms
- In The News
- (Corrections)
- Long Content
- Hard To Categorize

Sign up [here](#) or browse past newsletters [here](#). I'm considering creating a Patreon or substack for this newsletter; if you have any strong views, leave a comment.

Prediction Markets & Forecasting Platforms

Facebook's Forecast app now [out of beta](#) in the US and Canada.

Hypermind, a prediction market with virtual points but occasional monetary rewards, is organizing a [contest](#) for predicting US GDP in 2020, 2021 and 2022. Prizes sum up to \$90k.

Metaculus held the [Road to Recovery](#), and [20/20 Insight Forecasting](#) contests. It and collaborators also posted the results of their [2020 U.S. Election Risks Survey](#).

[CSET](#) publishes a report on using forecasters to inform big picture policy questions.

We illustrate Foretell's methodology with a concrete example: First, we describe three possible scenarios, or ways in which the tech-security landscape might develop over the next five years. Each scenario reflects different ways in which U.S.-China tensions and the fortunes of the artificial intelligence industry might develop. Then, we break each scenario down into near-term predictors and identify one or more metrics for each predictor. We then ask the crowd to forecast the metrics. Lastly, we compare the crowd's forecasts with projections based on historical data to identify trend departures: the extent to which the metrics are expected to depart from their historical trajectories.

Replication Markets opens their [Prediction Market for COVID-19 Preprints](#). Surveys opened on October 28, and markets will open on November 11, 2020.

In the News

The European Union is attempting to build a model of the Earth [at 1km resolution](#) as a test ground for its upcoming supercomputers. Typical models run at a resolution of 10 to 100km.

Michael Gove, a British Minister, gave a [66% chance to a Brexit deal](#). The Independent follows up by [giving the probabilities of different experts](#)

Some 538 highlights:

- US general election polls are generally a random walk, rather than [having momentum](#).
- [Pollsters have made some changes since 2016](#), most notably weighing by education.
- An [interactive presidential forecast](#)

[New York magazine](#) goes over some differences between 538's and The Economist's forecast for the US election.

[Reuters](#) looks at the volatility between the dollar and the yen or Swiss franc as a proxy for tumultuous elections. Reuters' interpretation is that a decline in long-run volatility implies that the election is not expected to be contested.

Meanwhile, new systems for [forecasting outbreaks](#) in the American pork industry may help prevent outbreaks, and also make the industry more profitable.

- On the topic of animals, see also a Metaculus question on whether [the EU will announce going cage-free by 2024](#).

Corrections

In the September newsletter, I claimed that bets on the order of \$50k could visibly move Betfair's odds. I got some [pushback](#). I asked Betfair itself, and their answer was:

It would definitely be an oversimplification to say that "markets can be moved with 10 to 50k", because it would depend on a number of other factors such as how much is available at that price at any one time and if anyone puts more money up at that price once all available money is taken.

For example if someone placed £100k on Biden at 1.44 and there was £35k at 1.45, and £57k at 1.44, then around £7k would be unmatched and the market would now be 1.43-1.44 on Biden. But if someone else still thinks the price should remain at 1.45-1.46 they could place bets to get it back to that, so the market will shift back almost immediately.

So to clarify, the bets outlined in those articles aren't necessarily the sole reason for the market moving, therefore they can't be deemed the causal connection. They are just headline examples to provide colour to the betting patterns at the time. I hope that is useful, let me know if you need any more info.

Negative Examples

Boeing [releases](#) an extremely positive market outlook. "A year ago, Boeing was predicting services market demand to be \$3.13 trillion from 2019-2028, making the

prediction for \$3 trillion from 2020-2029 look optimistic."

Long Content

The [World Agricultural Supply and Demand Estimates](#) is a monthly report by the US Department of Agriculture. It provides monthly estimates and past figures for crops worldwide, and for livestock production in the US specifically (meat, poultry, dairy), which might be of interest to the animal suffering movement. It also provides estimates of the past reliability of those forecasts. The October report can be found [here](#), along with a summary [here](#). The image below presents the 2020 and 2021 predictions, as well as the 2019 numbers:

U.S. Quarterly Animal Product Production 1/

Year and Quarter		Beef	Pork	Red Meat 2/	Broiler	Turkey	Total Poultry 3/	Red Meat & Poultry	Egg	Milk
2019	IV Annual	7,001 27,155	7,478 27,638	14,535 55,015	11,175 43,905	1,467 5,818	12,773 50,251	27,308 105,266	2,414 9,447	54.0 218.4
2020	I	6,929	7,426	14,407	11,237	1,469	12,841	27,248	2,349	56.1
	II	6,054	6,311	12,417	10,940	1,369	12,446	24,863	2,249	56.0
	III	7,105	7,060	14,215	11,275	1,450	12,866	27,081	2,310	55.2
	IV* Annual	7,050	7,360	14,466	11,150	1,450	12,735	27,201	2,345	54.9
	Sep Proj.	27,048	28,237	55,496	44,552	5,708	50,801	106,296	9,248	222.0
	Oct Proj.	27,138	28,157	55,505	44,602	5,738	50,888	106,392	9,253	222.3
2021	I* II* III* Annual	6,845 6,935 6,845	7,245 6,740 7,050	14,142 13,732 13,947	11,000 11,165 11,505	1,420 1,425 1,445	12,550 12,730 13,090	26,692 26,462 27,037	2,290 2,290 2,365	56.3 57.6 56.0
	Sep Proj.	27,355	28,445	56,018	45,020	5,770	51,330	107,348	9,375	225.4
	Oct Proj.	27,365	28,510	56,092	45,060	5,770	51,370	107,462	9,390	225.5

* Projection. 1/ Commercial production for red meats; federally inspected for poultry meats. 2/ Beef, pork, veal and lamb & mutton. 3/ Broilers, turkeys and mature chicken.

The Atlantic considers scenarios under which [Trump refuses to concede](#). Warning: very long, very chilling.

National Geographic on [the limits and recent history of weather forecasting](#). There are reasons to think that forecasting the weather accurately two weeks in advance might be difficult.

Andreas Stuhlmüller, of Ought, plays around with GPT-3 to [output probabilities](#); I'm curious to see what comes out of it. I'd previously tried (and failed) to get GPT-3 to output reasonable probabilities for Good Judgment Open questions.

A 2019 paper by Microsoft on [End-User Probabilistic Programming](#), that is, on adding features to spreadsheet software to support uncertain values, quantify uncertainty, propagate errors, etc.

The [2020 Presidential Election Forecasting symposium](#) presents 12 different election forecasts, ranging from blue wave to Trump win. [Here](#) is an overview.

[Blue Chip Economic Indicators](#) and [Blue Chip financial forecasts](#) are an extremely expensive forecasting option for various econometric variables. A monthly suscription costs \$2,401.00 and \$2,423.00, respectively, and provides forecasts by 50 members of prestigious institutions ("Survey participants such as Bank of America, Goldman Sachs & Co., Swiss Re, Loomis, Sayles & Company, and J.P. MorganChase, provide forecasts..."). An estimate of previous track record and accuracy [isn't available](#) before purchase. Further information on [Wikipedia](#)

- [Chief U.S. economist Ellen Zentner of Morgan Stanley](#) won the [Lawrence R. Klein Award](#) for the most accurate econometric forecasts among the 50 groups who participate in Blue Chip financial forecast surveys.
- I would be very curious to see if Metaculus' top forecasters, or another group of expert forecasts, could beat the Blue Chips. I'd also be curious how they fared on January, February and March of this year.

Hard to categorize.

Scientists use [precariously balanced rock formations to improve accuracy of earthquake forecasts](#). They can estimate when the rock formation appeared, and can calculate what magnitude an earthquake would have had to be to destabilize it. Overall, a neat proxy.

Some [superforecasters](#) to follow on twitter.

Dart Throwing Spider Monkey proudly presents [Intro to Forecasting 01 - What is it and why should I care?](#) and [Intro to Forecasting 02 - Reference class forecasting](#).

I've gone through the Effective Altruism Forum and LessWrong and added or made sure that the forecasting tag is applied to the relevant posts for October (LessWrong [link](#), Effective Altruism forum [link](#)). This provides a change-log for the month. For the Effective Altruism forum, this only includes Linch Zhang's post on [Some learnings I had from forecasting in 2020](#). For LessWrong, this also includes a [post announcing that Forecast, a prediction platform by Facebook](#) is now out of beta.

Note to the future: All links are added automatically to the Internet Archive. In case of link rot, go [there](#) and input the dead link.

Using actuarial life tables and an adjustment for covid, the implied probability that all 246 readers of this newsletter drop dead before the next month is at least 10^{-900} (if they were uncorrelated). See [this Wikipedia page](#) or [this xkcd comic](#) for a comparison with other low probability events, such as asteroid impacts.

Forecasting Newsletter: November 2020

Highlights

- DeepMind claims a major [breakthrough](#) in protein folding.
- OPEC forecasts [slower growth](#)
- Gnosis announces [futarchy experiment](#)

Index

- Highlights
- In The News
- Prediction Markets & Forecasting Platforms
- United States Presidential Elections Post-mortems
- Hard To Categorize
- Long Content

Sign up [here](#) or browse past newsletters [here](#).

In the News

DeepMind claims a major breakthrough in protein folding ([press release](#), [secondary source](#))

DeepMind has developed a piece of AI software called AlphaFold that can accurately predict the structure that proteins will fold into in a matter of days.

This computational work represents a stunning advance on the protein-folding problem, a 50-year-old grand challenge in biology. It has occurred decades before many people in the field would have predicted. It will be exciting to see the many ways in which it will fundamentally change biological research.

Figuring out what shapes proteins fold into is known as the "protein folding problem", and has stood as a grand challenge in biology for the past 50 years. In a major scientific advance, the latest version of our AI system AlphaFold has been recognised as a solution to this grand challenge by the organisers of the biennial Critical Assessment of protein Structure Prediction (CASP). This breakthrough demonstrates the impact AI can have on scientific discovery and its potential to dramatically accelerate progress in some of the most fundamental fields that explain and shape our world.

In the results from the 14th CASP assessment, released today, our latest AlphaFold system achieves a median score of 92.4 GDT overall across all targets. This means that our predictions have an average error (RMSD) of approximately 1.6 Angstroms, which is comparable to the width of an atom (or 0.1 of a nanometer). Even for the very hardest protein targets, those in the most

challenging free-modelling category, AlphaFold achieves a median score of 87.0 GDT.

Crucially, CASP chooses protein structures that have only very recently been experimentally determined (some were still awaiting determination at the time of the assessment) to be targets for teams to test their structure prediction methods against; they are not published in advance. Participants must blindly predict the structure of the proteins.

The Organization of the Petroleum Exporting Countries (OPEC) forecasts slower growth and slower growth in oil demand ([primary source](#), [secondary source](#).) In particular, it forecasts long-term growth for OECD countries — which I take to mean that growth because of covid recovery is not counted — to be below 1%. On the one hand, their methodology is opaque, but on the other hand, I expect them to actually be trying to forecast growth and oil demand, because it directly impacts the amount of barrels it is optimal for them to produce.

Google and Harvard's Global Health Institute update their US covid model, and publish it on NeurIPS 2020 ([press release](#)), aiming to be robust, interpretable, extendable, and to have longer time horizons. They're also using it to advertise various Google products. It has been extended to [Japan](#).

Prediction Markets & Forecasting Platforms

Gnosis announces the GnosisDAO ([announcement](#), [secondary source](#)), an organization governed by prediction markets (i.e., a [futarchy](#)): "The mission of GnosisDAO is to successfully steward the Gnosis ecosystem through futarchy: governance by prediction markets."

Metaculus have a new report on forecasting covid vaccines, testing and economic impact ([summary](#), [full report](#)). They also organized [moderator elections](#) and are hiring for a [product manager](#).

Prediction markets have kept selling Trump not to be president in February at \$0.85 to \$0.9 (\$0.9 as of now, where the contract resolves to \$1 if Trump isn't president in February.) Non-American readers might want to explore [PolyMarket](#) or [FTX](#), American readers with some time on their hands might want to actually put some money into [PredictIt](#). Otherwise, some members of the broader Effective Altruism and rationality communities made a fair amount of money betting on the election.

CSET recorded [Using Crowd Forecasts to Inform Policy](#) with [Jason Matheny](#), CSET's Founding Director, previously Director of IARPA. I particularly enjoyed the verbal history bits, the sheer expertise Jason Matheny radiated, and the comments on how the US government currently makes decisions.

Q: Has the CIA changed its approach to using numbers rather than words?

A: No, not really. They use some prediction markets, but most analytic products are still based on verbiage.

As a personal highlight, I was referred to as "top forecaster Sempere" towards the end of [this piece](#) by CSET. I've since then lost the top spot, and I'm back to holding the second place.

I also organized the [Forecasting Innovation Prize \(LessWrong link\)](#), which offers \$1000 for research and projects on judgemental forecasting. For inspiration, see the [project suggestions](#). Another post of mine, [Predicting the Value of Small Altruistic Projects: A Proof of Concept Experiment](#) might also be of interest to readers in the Effective Altruism community. In particular, I'm looking for volunteers to expand it.

Negative Examples

[Release of Covid-19 second wave death forecasting 'not in public interest', claims Scottish Government](#)

The Scottish Government has been accused of "absurd" decision making after officials blocked the release of forecasting analysis examining the potential number of deaths from a second wave of Covid-19.

Officials refused to release the information on the basis that it related to the formulation or development of government policy and was "not in the public interest" as it could lead to officials not giving "full and frank advice" to ministers.

The response also showed no forecasting analysis had been undertaken by the Scottish Government over the summer on the potential of a second wave of Covid-19 on various sectors.

United States Presidential Election Post-mortems

Thanks to the Metaculus Discord for suggestions for this section.

Independent postmortems

- David Glidden's ([@dglid](#)) comprehensive [spreadsheet](#) comparing 538, the Economist, Smarts and PredictIt in terms of Brier scores for everything. tl;dr: Prediction Markets did better in closer states. (see [here](#) for the log score.)
- [Hindsight is 2020](#); a nuanced take.
- [2020 Election: Prediction Markets versus Polling/Modeling Assessment and Postmortem](#).

"We find a market that treated day after day of good things for Biden and bad things for Trump, in a world in which Trump was already the underdog, as not relevant to the probability that Trump would win the election."

Markets overreacted during election night.

[On methodology:] You bet into the market, but the market also gets to bet into your fair values. That makes it a fair fight." [Note: see [here](#) for a graph through time, and [here](#) for the orginal, though less readable source]

...polls are being evaluated, as I've emphasized throughout, against a polls plus humans hybrid. They are not being evaluated against people who don't

look at polls. That's not a fair comparison.

- [Partisans, Sharps, And The Uninformed Quake US Election Market](#). tl;dr: "I find myself really torn between wanting people to be more rational and make better decisions. And then also, like, well, I want people to offer 8-1 on Trump being in office in February."

American Mainstream Media

Mostly unnuanced.

- The Cook Political Report on [Why Couldn't Democrats Ride the Blue Wave?](#). tl;dr: "If you wanted to sum up the election results in a few words, those words might be that by the barest of majority, voters were anti-Trump—but they were not anti-Republican."
- Wall Street Journal's [The Price of Bad Polling \(unpaywalled archive link\)](#)
- ABC news: [Were 2020 election polls wrong?](#)
- The New York Times: [What Went Wrong With Polling? Some Early Theories](#) and [Why Political Polling Missed the Mark. Again.](#)
- Fox News (handpicked for interestingness; see [here](#) for more representative sample): [Stock market predicts Trump will defeat Biden](#); [Nate Silver defends his analysis of 2020 election polls](#); [Frank Luntz urges pollsters to seek new profession after Trump outperforms polls: 'Sell real estate'](#); [Karl Rove says Trump outperforming polls was 'remarkable achievement'](#).

FiveThirtyEight.

- [We Have A Lot Of New Polls, But There's Little Sign Of The Presidential Race Tightening](#)
- [FiveThirtyEight's Final 2020 presidential election forecast](#)
- [Biden Won — Pretty Convincingly In The End](#)

Andrew Gelman.

- [Don't kid yourself. The polls messed up—and that would be the case even if we'd forecasted Biden losing Florida and only barely winning the electoral college](#)
- [Comparing election outcomes to our forecast and to the previous election](#)

As we've discussed elsewhere, we can't be sure why the polls were off by so much, but our guess is a mix of differential nonresponse (Republicans being less likely than Democrats to answer, even after adjusting for demographics and previous vote) and differential turnout arising from on-the-ground voter registration and mobilization by Republicans (not matched by Democrats because of the coronavirus) and maybe Republicans being more motivated to go vote on election day in response to reports of 100 million early votes.

- [So, what's with that claim that Biden has a 96% chance of winning?](#)
- [Why we are better off having election forecasts](#)
- See also: [Multilevel regression with poststratification](#).

Hard to Categorize

[Forbes on how to improve hurricane forecasting:](#)

...to greatly improve the hurricane intensity forecast, we need to increase the subsurface ocean measurements by at least one order of magnitude...

One of the most ambitious efforts to gather subsurface data is Argo, an international program designed to build a global network of 4,000 free-floating sensors that gather information like temperature, salinity and current velocity in the upper 2,000 meters of the ocean.

Argo is managed by NOAA's climate office that monitors ocean warming in response to climate change. This office has a fixed annual budget to accomplish the Argo mission. The additional cost of expanding Argo's data collection by 10 times doesn't necessarily help this office accomplish the Argo mission. However, it would greatly improve the accuracy of hurricane forecasts, which would benefit the NOAA's weather office — a different part of NOAA. And the overall benefit of improving even one major hurricane forecast would be to save billions [in economic losses], easily offsetting the entire cost to expand the Argo mission.

[In wake of bad salmon season, Russia calls for new forecasting approach:](#)

In late October, Ilya Shestakov, head of the Russian Federal Agency for Fisheries, met with Russian scientists from the Russian Research Institute of Fisheries and Oceanography (VNIRO) to talk about the possible reasons for the difference. According to scientists, the biggest surprises came from climate change.

"We have succeeded in doing a deeper analysis of salmon by the combination of fisheries and academic knowledge added by data from longstanding surveys," Marchenko said. "No doubt, we will be able to enhance the accuracy of our forecasts by including climate parameters into our models."

[Political Polarization and Expected Economic Outcomes \(summary\)](#)

"87% of Democrats expect Biden to win while 84% of Republicans expect Trump to win"

"Republicans expect a fairly rosy economic scenario if Trump is elected but a very dire one if Biden wins. Democrats ... expect calamity if Trump is re-elected but an economic boom if Biden wins."

Dart Throwing Spider Monkey proudly presents the third part of his Intro to Forecasting series: [Building Probabilistic Intuition](#)

[A gentle introduction to information charts](#): a simple tool for thinking about probabilities in general, but in particular for predictions with a sample size of one.

[A youtube playlist with forecasting content](#) h/t Michal Dubrawski.

[Farm-level outbreak forecasting tool expands to new regions](#)

An article with some examples of [Crime Location Forecasting](#), and on whether it can be construed as entrapment.

[Why Forecasting Snow Is So Difficult](#): Because it is very sensitive to initial conditions.

[Google looking for new ways to predict cyber-attackers' behavior.](#)

Long Content

[Taking a disagreeing perspective improves the accuracy of people's quantitative estimates](#), but this depends on the question type.

...research suggests that the same principles underlying the wisdom of the crowd also apply when aggregating multiple estimates from the same person – a phenomenon known as the "wisdom of the inner crowd"

Here, we propose the following strategy: combine people's first estimate with their second estimate made from the perspective of a person they often disagree with. In five pre-registered experiments (total N = 6425, with more than 53,000 estimates), we find that such a strategy produces highly accurate inner crowds (as compared to when people simply make a second guess, or when a second estimate is made from the perspective of someone they often agree with). In explaining its accuracy, we find that taking a disagreeing perspective prompts people to consider and adopt second estimates they normally would not consider as viable option, resulting in first- and second estimates that are highly diverse (and by extension more accurate when aggregated). However, this strategy backfires in situations where second estimates are likely to be made in the wrong direction. Our results suggest that disagreement, often highlighted for its negative impact, can be a powerful tool in producing accurate judgments.

..after making an initial estimate, people can be instructed to base their additional estimate on different assumptions or pieces of information. A demonstrated way to do this has been through "dialectical bootstrapping" where, when making a second estimate, people are prompted to question the accuracy of their initial estimate. This strategy has been shown to increase the accuracy of the inner crowd by getting the same person to generate more diverse estimates and errors.... as a viable method to obtain more diverse estimates, we propose to combine people's initial estimate with their second estimate made from the perspective of a person they often disagree with.... although generally undesirable, research in group decision-making indicates that disagreement between individuals may actually be beneficial when groups address complex problems. For example, groups consisting of members with opposing views and opinions tend to produce more innovative solutions, while polarized editorial teams on Wikipedia (i.e., teams consisting of ideologically diverse sets of editors) produce higher quality articles...

These effects occur due to the notion that disagreeing individuals tend to produce more diverse estimates, and by extension errors, which are cancelled out across group members when averaged. ...we conducted two (pre-registered) experiments...

People who made a second estimate from the perspective of a person they often disagree with benefited more from averaging than people who simply made a second guess.

... However, although generally beneficial, this strategy backfired in situations where second estimates were likely to be made in the wrong direction. [...] For example, imagine being asked the following question: "What percent of China's population identifies as Christian?". The true answer to this question is 5.1% and if you are like most people, your first estimate is probably leaning towards this lower

end of the scale (say your first estimate is 10%). Given the position of the question's true answer and your first estimate, your second estimate is likely to move away from the true answer towards the opposite side of the scale (similar to the scale-end-effect⁴⁵), effectively hurting the accuracy of the inner crowd.

We predicted that the average of two estimates would not lead to an accuracy gain in situations where second estimates are likely to be made in the wrong direction. We found this to be the case when the answer to a question was close to the scale's end (e.g., an answer being 2% or 98% on a 0%-100% scale).

[A 2016 article attacking Nate Silver's model](#), key to understanding why Nate Silver is often so smug.

[Historical Presidential Betting Markets](#), in the US before 2004.

...we show that the market did a remarkable job forecasting elections in an era before scientific polling. In only one case did the candidate clearly favored in the betting a month before Election Day lose, and even state-specific forecasts were quite accurate. This performance compares favorably with that of the Iowa Electronic Market (currently [in 2004] the only legal venue for election betting in the United States). Second, the market was fairly efficient, despite the limited information of participants and attempts to manipulate the odds by political parties and newspapers. The extent of activity in the presidential betting markets of this time was astonishingly large. For brief periods, betting on political outcomes at the CurbExchange in New York would exceed trading in stocks and bonds.

Covering developments in the Wall Street betting market was a staple of election reporting before World War II. Prior to the innovative polling efforts of Gallup, Roper and Crossley, the other information available about future election outcomes was limited to the results from early-season contests, overtly partisan canvasses and straw polls of unrepresentative and typically small samples. The largest and best-known nonscientific survey was the Literary Digest poll, which tabulated millions of returned postcard ballots that were mass mailed to a sample drawn from telephone directories and automobile registries. After predicting the presidential elections correctly from 1916 to 1932, the Digest famously called the 1936 contest for Landon in the election that F. Roosevelt won by the largest Electoral College landslide of all time. Notably, although the Democrat's odds prices were relatively low in 1936, the betting market did pick the winner correctly. The betting quotes filled the demand for accurate odds from a public widely interested in wagering on elections. In this age before mass communication technologies reached into America's living rooms, election nights were highly social events, comparable to New Year's Eve or major football games. In large cities, crowds filled restaurants, hotels and sidewalks in downtown areas where newspapers and brokerage houses would publicize the latest returns and people with sporting inclinations would wager on the outcomes. Even for those who could not afford large stakes, betting in the run-up to elections was a cherished ritual. A widely held value was that one should be prepared to "back one's beliefs" either with money or more creative dares. Making freak bets—where the losing bettor literally ate crow, pushed the winner around in a wheelbarrow or engaged in similar public displays—was wildly popular.

Gilliams (1901, p. 186) offered "a moderate estimate" that in the 1900 election "there were fully a half-million such [freak]bets—about one for every thirty

voters." In this environment, it is hardly surprising that the leading newspapers kept their readership well informed about the latest market odds.

The newspapers recorded many betting and bluffing contests between Col. Thomas Swords, Sergeant of Arms of the National Republican Party, and Democratic betting agents representing Richard Croker, Boss of Tammany Hall, among others. In most but not all instances, these officials appear to bet in favor of their party's candidate; in the few cases where they took the other side, it was typically to hedge earlier bets.

...In conclusion, the historical betting markets do not meet all of the exacting conditions for efficiency, but the deviations were not usually large enough to generate consistently profitable betting strategies using public information

The newspapers reported substantially less betting activity in specific contests and especially after 1940. In part, this reduction in reporting reflected a growing reluctance of newspapers to give publicity to activities that many considered unethical. There were frequent complaints that election betting was immoral and contrary to republican values. Among the issues that critics raised were moral hazard, election tampering, information withholding and strategic manipulation.

In response to such concerns, New York state laws did increasingly attempt to limit organized election betting. Casual bets between private individuals always remained legal in New York. However, even an otherwise legal private bet on elections technically disqualified the participants from voting—although this provision was rarely enforced—and the legal system also discouraged using the courts to collect gambling debts. Anti-gambling laws passed in New York during the late 1870s and the late 1900s appear to put a damper on election betting, but in both cases, the market bounced back after the energy of the moral reformers flagged. Ultimately, New York's legalization of parimutuel betting on horse races in 1939 may have done more to reduce election betting than any anti-gambling policing. With horseracing, individuals interested in gambling could wager on several contests promising immediate rewards each day, rather than waiting through one long political contest.

The New York Stock Exchange and the CurbMarket also periodically tried to crack down. The exchanges characteristically did not like the public to associate their socially productive risk-sharing and risk-taking functions with gambling on inherently zero-sum public or sporting events. In the 1910s and again after the mid-1920s, the stock exchanges passed regulations to reduce the public involvement of their members. In May 1924, for example, both the New York Stock Exchange and the Curb Market passed resolutions expressly barring their members from engaging in election gambling. After that, while betting activity continued to be reported in the newspapers, the articles rarely named the participants. During the 1930s, the press noted that securities of private electrical utilities had effectively become wagers on Roosevelt (on the grounds that New Deal policy initiatives such as the formation of the Securities and Exchange Commission and the Tennessee Valley Authority constrained the profits of existing private utilities).

A final force pushing election betting underground was the rise of scientific polling. For newspapers, one of the functions of reporting Wall Street betting odds had been to provide the best available aggregate information [...] The scientific

polls, available on a weekly basis, provided the media with a ready substitute for the betting odds, one not subject to the moral objections against gambling.

In summer 2003, word leaked out that the Department of Defense was considering setting up a Policy Analysis Market, somewhat similar to the Iowa Electronic Market, which would seek to provide a market consensus about the likelihood of international political developments, especially in the Middle East. Critics argued that this market was subject to manipulation by insiders and might allow extremists to profit financially from their actions.

Note to the future: All links are added automatically to the Internet Archive. In case of link rot, go [there](#) and input the dead link.

"I'd rather be a bookie than a goddamned poet." — Sherman Kent, 1964, when pushing for more probabilistic forecasts and being accused of trying to turn the CIA into "the biggest bookie shop in town."

Forecasting Newsletter: December 2020

Highlights

- Nigh unbeatable forecaster gives [85% chance that the newly identified COVID-19 strain is >30% more transmissible](#).
- Prediction markets and betting platforms mostly resolved the election in favor of Biden already. However, new markets have been created about whether Trump will still be president in February. Those who believe he won't can still earn a circa [10% return](#).
- Metaculus announces their [AI Progress tournament](#), with \$50,000 in rewards.

Index

- Prediction Markets & Forecasting Platforms
- US Presidential Election Betting
- In the News
- Negative Examples
- Hard to Categorize
- Long Content
- Yearly Housekeeping

Sign-up or view past newsletters [here](#).

Prediction Markets & Forecasting Platforms

Metaculus organized the [AI Progress tournament](#), covered by a Forbes contributor [here](#); rewards are hefty (\$50,000 in total). Questions for the first round, which focuses on the state of AI exactly six months into the future, can be found [here](#). In the [discussion page](#) for the first round, some commenters point out that the questions so far aren't that informative or intellectually stimulating. Metaculus has also [partnered with The Economist](#) for a series of events in 2021, and were mentioned in this [article](#) (sadly paywalled). They are also [hiring](#).

[@lxrjl](#), a moderator in the platform, has gathered a [list of forecasts about Effective Altruism organisations which are currently on Metaculus](#). This includes a [series](#) introducing ACE (Animal Charity Evaluators) to using forecasting as a tool to inform their strategy, which Misha Yagudin and I came up with after some back-and-forth with ACE.

Augur partnered with the crypto currency/protocol \$COVER/CoverProtocol to provide protection from losses in case Augur got hacked. In effect, traders could also bet for or against the proposition that a given Augur market will be hacked, and they could do this in a separate protocol. Traders could then use these bets to provide or acquire insurance ([source](#), [secondary source](#)).

In an ironic twist of fate, soon after commencing the partnership CoverProtocol was [hacked](#) ([details](#), [secondary source](#)). Even though the money was later [returned](#), it seems that \$COVER is being delisted from major cryptocurrency trading exchanges.

I added [catnip.exchange](#) to [this](#) list of prediction markets Jacob Lagerros created. More suggestions are welcome, and can be made by leaving a comment in the linked document.

During Christmas, I programmed an [interface to view PolyMarket](#) markets and trades, using their GraphQL API. PolyMarket is a speculative crypto prediction market, which I trust because they successfully resolved the first round of US presidential elections without

absconding with the money. PolyMarket's frontpage has an annoying UI bug— it sometimes doesn't show the 2% liquidity provider fee, and my interface solves that.

Karen Hagar, with the collaboration of Scott Eastman, has started two **forecasting-related nonprofit** organizations:

[AZUL Foresight](#) is devoted to forecasting and red team analysis, and is planning to have a geopolitical analysis column.

[LogicCurve](#) is devoted to forecasting education, training and international outreach.

Both Karen and Scott are Superforecasters™ and “friends of the newsletter”; we previously worked together on EpidemicForecasting predicting the spread of COVID-19 in developing nations. They are eager to get started on forecasting, and are looking for clients. I'm curious to see what comes of it, particularly because forecasting can be combined with almost any interesting and important problem.

US Presidential Election Betting

The primary story for prediction markets this month was that they generally acknowledged Biden as the president of the USA, early. This was done according to the terms and conditions of the markets—they were to be resolved according to whomever Fox, CNN, AP, and other major American news outlets called as the winner. However, the resolution was also seen as problematic by those who believe that Trump still has a chance. Prediction markets reacted by opening new questions asking whether Trump will still be president by February 2021. For example, [here](#) is one on FTX, and [here](#) is a similar one about Biden on PolyMarket.

As of late December one can still get, for example, a circa 1.1x return on PolyMarket by betting on Biden if he wins, or a circa 15x return by betting on FTX by betting on Trump, if Trump wins. FTX has Trump at ca. 6%, whereas PolyMarket has Biden at ca. 88%, and arbitrage hasn't leveled them yet (but note that their resolution criteria are different). However, the process of betting on one's preferred candidate depends on one's jurisdiction. As far as I understand (and I give ~50% to one of these being wrong or sub-optimal):

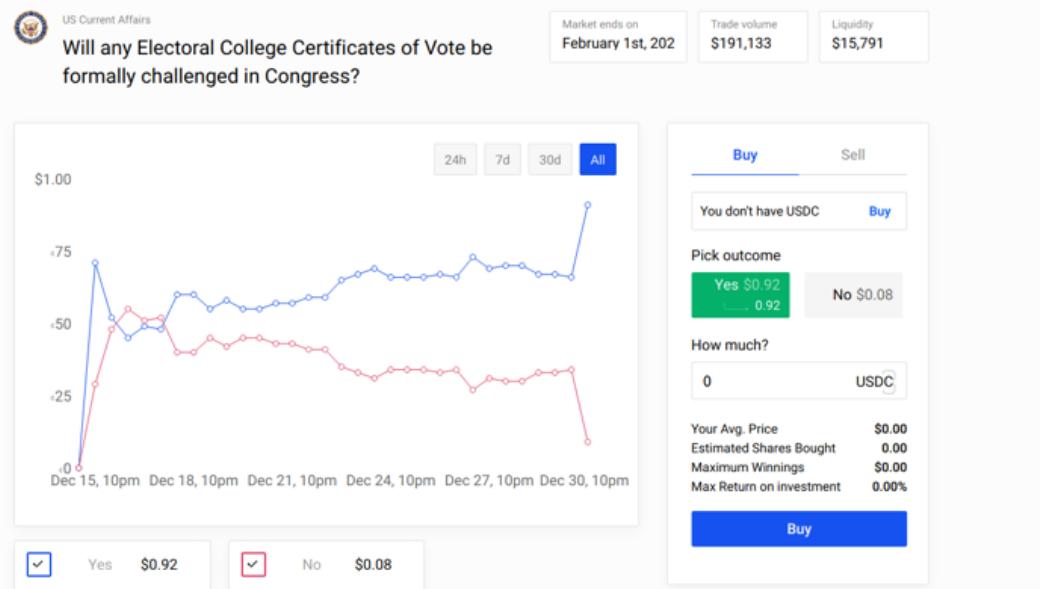
- Risk-averse American living in America: Use PredictIt.
- European, Russian, American living abroad, risk-loving American: Use FTX/PolyMarket.
- British: Use Smarts/PolyMarket/PaddyPower.

Other platforms that I haven't looked much into are catnip.exchange, Augur, and Omen. For some of these platforms, one needs to acquire USDC, a cryptocurrency pegged to the dollar. For this, I've been using crypto.com, but it's possible that Coinbase or other exchanges offer better rates. If you're interested in making a bet, you should do so before the 6th of January, one of the last checkpoints in the American election certification process. Note that getting your funds into a market might take a couple of days.

As for the object-level commentary, [here](#) is a piece by Vox, [here](#) one by the New York Times, and [here](#) one by The Hill. With regards to the case for Trump staying in the White House, [here](#) is a Twitter thread collecting information from a vocal member of the PolyMarket Discord Server. A selection of resources from that thread is:

- [This site](#) aims to collect all instances of purported election fraud and manipulation in the US.
- [This thread](#) elaborates on the role which Trump supporters are hoping Vice President Pence will take during the 6th of January joint session of the US Congress.
- An [ELI5 on how Trump could win](#).

I'm giving more space to the views I disagree with, because I'm betting some money that Biden will, in fact, be inaugurated president, though nothing I can't afford to lose. I'm also aware that [in matters of politics](#), it's particularly easy to confuse a 30% for a 3% chance, so I wouldn't recommend full [Kelly betting](#).



As an interesting tidbit, most big bets on PolyMarket's election markets are against Trump. For instance, the largest bet placed against Trump amounts to \$1,480,000, whereas the largest bet placed on his success is \$622,223.

In the News

A [new variant of COVID-19](#) has been identified (New York Times coverage [here](#)). The broader Effective Altruism and rationality communities are giving high probabilities to the possibility that this new variant is significantly more contagious. See [here](#), [here](#), [here](#), and the last section [here](#). Juan Cambeiro, an expert COVID-19 forecaster who has consistently outperformed competitors and experts across Metaculus and various Good Judgement platforms, gives an 85% chance that this specific new strain is >30% more transmissible. Because of Cambeiro's past forecasting prowess, subject-matter expertise and nigh unbeatable track record, I suggest adopting his probabilities as your own, and acting accordingly. Note that the Metaculus questions he is forecasting are clearly, precisely, narrowly and tightly defined, so there isn't room for doubt in that regard.



Coronavirus Outbreak

CLOSING
Jan 1, 2021 09:00AM (11 hours)

The coronavirus outbreak poses challenges for the economies, foreign policies, and societies of the international community as a whole, as well as public-health concerns for those infected or at risk. Help to forecast how the evolution of this disease will shape the world order.

Questions	Leaderboard	Activity			
Rank	Username	No. of Questions Forecasted	Brier	Median	Relative Brier Score ⓘ
1	 juancambeiro	63	0.188	0.26	-3.839
2	 Dillyweed	64	0.234	0.294	-3.226
3	 jainsa	44	0.204	0.302	-3.196
4	 WBHumanoid	53	0.207	0.279	-3.068
5	 ThWedler	34	0.286	0.37	-2.586
6	 Leonardthethird	36	0.224	0.312	-2.49
7	 K-for-clean-H2O	8	0.151	0.45	-2.292
8	 cdob63	70	0.27	0.302	-2.245
9	 Dogopotamus	34	0.231	0.304	-2.089
10	 Jesse-Harris	41	0.223	0.252	-2.038

Budget forecasting in the US under COVID-19:

- [Kentucky decides to go with a forecast which is very conservative and doesn't incorporate all available information](#). This is an interesting example of a legitimate use of forecasting which doesn't involve maximizing accuracy. The forecast will be used to craft the budget, so choosing a more pessimistic forecast might make sense if one is aiming for increased robustness. Or, in other words, the forecast isn't aiming to predict the expected revenue, but rather the lower bound of an 80% confidence interval.
- [Louisiana](#) chooses to delay their projections until early next year.
- [Colorado](#) finds that they have a \$3.75 billion surplus on a \$32.5 billion budget after budget cuts earlier in the year.

The Washington post editorializes: [Maverick astrophysicist calls for unusually intense solar cycle, straying from consensus view \(original source\)](#). On this topic, see also the [New Solar Cycle 25 Question Series](#) on Metaculus.

Our method predicts that SC25 [the upcoming sunspot cycle] could be among the strongest sunspot cycles ever observed, depending on when the upcoming termination happens, and it is highly likely that it will certainly be stronger than present SC24 (sunspot number of 116) and most likely stronger than the previous SC23 (sunspot number of 180). This is in stark contrast to the consensus of the SC25PP, sunspot number maximum between 95 and 130, i.e. similar to that of SC24.

An opinion piece by [The Wall Street Journal](#) talks about measures taken by the US Geological Survey to make climate forecasts less political ([unpaywalled archive link](#)).

The approach includes evaluating the full range of projected climate outcomes, making available the data used in developing forecasts, describing the level of uncertainty in the findings, and periodically assessing past expectations against actual performance to provide guidance on future projections.

Moving forward, this logical approach will be used by the USGS and the Interior Department for all climate-related analysis and research—a significant advancement in the government's use and presentation of climate science.

These requirements may seem like common sense, but there has been wide latitude in how climate assessments have been used in the past. This new approach will improve scientific efficacy and provide a higher degree of confidence for policy makers responding to potential future climate change conditions because a full range of plausible outcomes will be considered.

Science should never be political. We shouldn't treat the most extreme forecasts as an inevitable future apocalypse. The full array of forecasts of climate models should be considered. That's what the USGS will do in managing access of natural resources and conserving our natural heritage for the American people.

Australian weather forecasters are [incorporating climate change comments](#) into their coverage ([archive link](#)).

As the year comes to a close, various news media are taking stock of past predictions, and making new predictions for 2021. These aren't numerical predictions, and as such are difficult to score. Some examples:

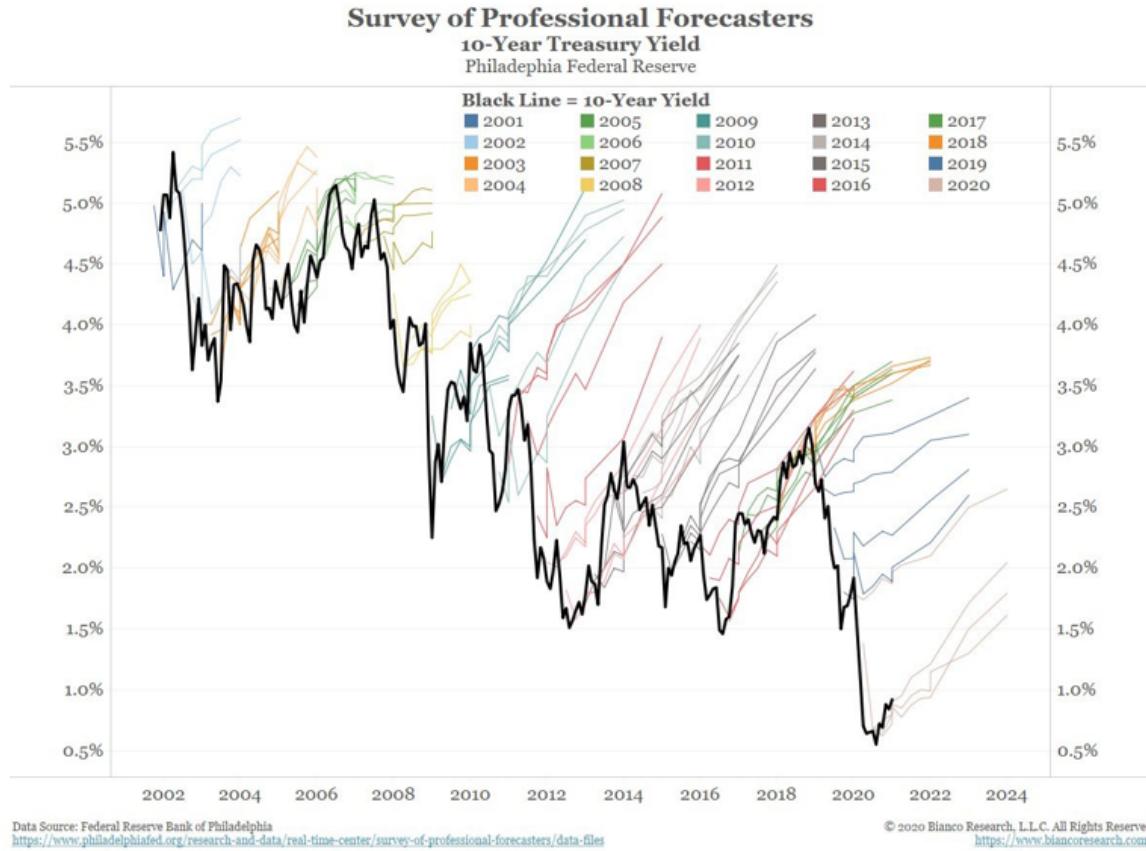
- Politico: [The Worst Predictions of 2020](#)
- New Statesman: [In January, I made ten predictions for 2020 – how did they turn out?](#)
- CryptoBriefing.com: [Crypto Predictions for 2020: Who Got It Right?](#)
- New York Times: [Clueless About 2020, Wall Street Forecasters Are at It Again for 2021](#)
- The Wall Street Journal: [Here's a Market Forecast: 2021 Will Be Hard to Predict](#)
- Financial Times: [Forecasting the world in 2021.](#)
- Forbes: [What Will The Stock Market Return In 2021?](#)
- Bloomberg: [Ignore All 2021 Market Predictions - Except This One](#)
- Washington Post: [Five \(somewhat\) upbeat predictions for 2021](#)
- News.Bitcoin.com: [Zero to \\$318,000: Proponents and Detractors Give a Variety of Bitcoin Price Predictions for 2021](#)

Negative Examples

Trump [predicted](#) that the US stock market would crash if Biden won. Though it still could, the forecast is not looking good. Here is CNN [making that point](#).

[Betting markets predicted no-deal after the failed Brexit summit](#). Though they did see a bump, the prediction market quoted [went up](#) briefly afterwards.

Twitter user [@kjhealy](#) visualizes forecasts from the [Survey of Professional Forecasters](#):



Hard to Categorize

Twitch adds [prediction functionalities](#) h/t [@Pongo](#).

The US's National Oceanic and Atmospheric Administration (NOAA) is organizing a [contest](#) to forecast movements in the Earth's magnetic field, with prizes totaling \$30,000.

The efficient transfer of energy from solar wind into the Earth's magnetic field causes geomagnetic storms. The resulting variations in the magnetic field increase errors in magnetic navigation. The disturbance-storm-time index, or Dst, is a measure of the severity of the geomagnetic storm.

In this challenge, your task is to develop models for forecasting Dst that push the boundary of predictive performance, under operationally viable constraints, using the real-time solar-wind (RTSW) data feeds from NOAA's DSCOVR and NASA's ACE satellites. Improved models can provide more advanced warning of geomagnetic storms and reduce errors in magnetic navigation systems.

The [US Congress adopts a plan to consolidate weather catastrophe forecasting](#). Previously, different agencies had been in charge of predicting weather phenomena.

A [new flood forecasting platform implemented in Guyana](#). As is becoming usual for these kinds of projects, I am unable to evaluate the extent to which this platform will be useful.

Despite Dominica and Guyana's agriculture sectors being the primary industries, the sector has constantly been affected by disasters. Recurring hurricanes, floods and droughts represent real threat to development and food security at the national level. It

also increases the vulnerability of local communities and puts small farmers, live stock holders, and aggro-processors, who are primarily women, at risk.

The [SlateStarCodex subreddit](#) talks about a [paper](#) by Andrew Gelman, which discusses flaws with "pure Bayesianism".

Bayesian updating only works if the "true model" is in the space of models you're updating over. This is never the case in practice. And, in fact Bayesian updating can lead you to becoming ever more convinced of a given model that is clearly false.

DartThrowingSpiderMonkey ([@alexjl](#)) presents the fourth video in his [Introduction to Forecasting](#) Series. This time it's about making Guesstimate models for questions for which a base rate is nonexistent or hard to find.

[xkcd](#) has a list of comparisons to help visualize different probabilities.

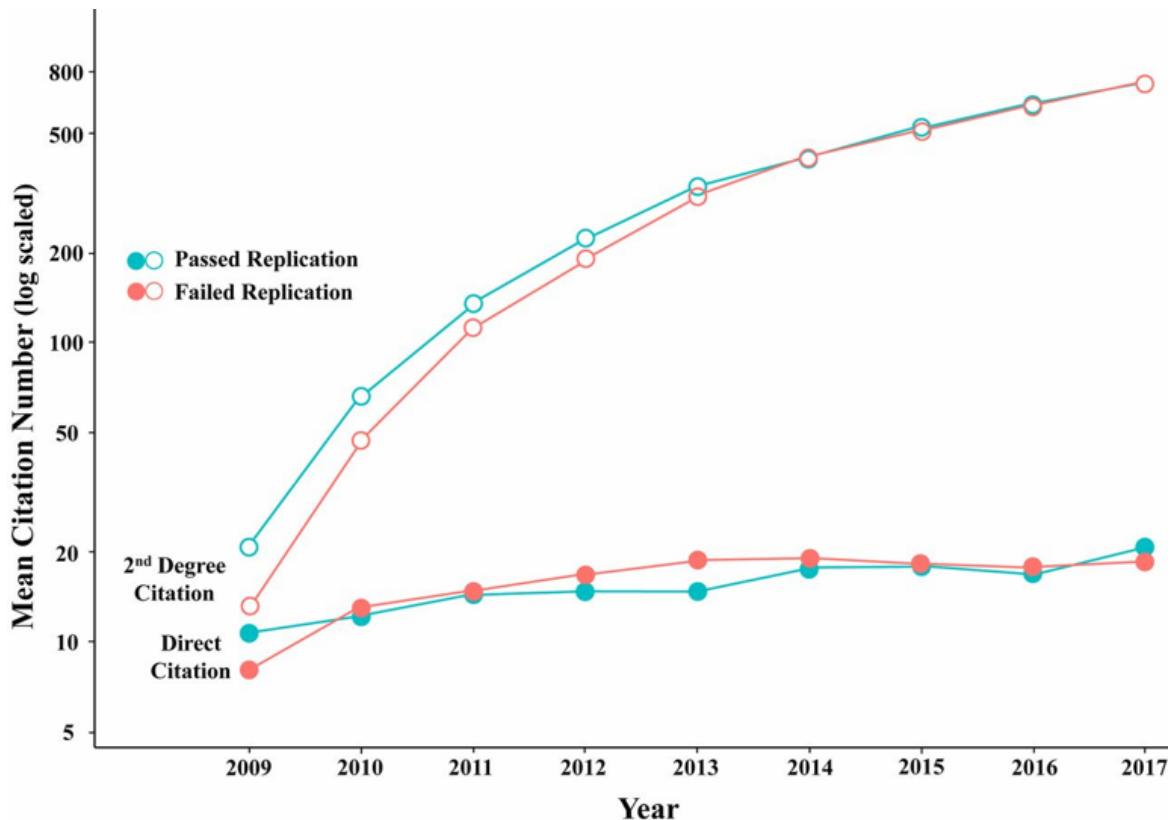
[Volcano forecasting models](#) might help New Zealand tourists who want to visit risky places.

Long Content

Paper: [Estimating the deep replicability of scientific findings using human and artificial intelligence \(secondary source\)](#).

The authors of the paper train a machine learning model to predict replicability of research results, using a relatively meager sample of 96 papers. The papers, taken from the original [Reproducibility Project: Psychology](#), have been validated on various other datasets. While the model does only slightly better than human prediction markets, it's important to note that a machine learning system, once set-up, would be much faster. Interestingly, the authors "did not detect statistical evidence of model bias regarding authorship prestige, sex of authors, discipline, journal, specific words, or subjective probabilities/persuasive language". Otherwise, their setup is relatively simple: a pre-processing step which translates words to vectors using [Word2vec](#), followed by a random forest combined with [bagging](#).

A nice tidbit from the paper is that past citation count isn't very predictive of future replication:



Article: [Forecasting the next COVID-19](#).

Implementing better measures and institutions to predict pandemics is probably a good idea. However, I'd expect the next catastrophic event in the scale of COVID-19 to [not be a pandemic](#).

Princeton disease ecologist C. Jessica Metcalf and Harvard physician and epidemiologist Michael Mina say that predicting disease could become as commonplace as predicting the weather. The Global Immunological Observatory, like a weather center forecasting a tornado or hurricane, would alert the world, earlier than ever before, to dangerous emerging pathogens like SARS-CoV-2.

A GIO [Global Immunological Observatory] would require an unprecedented level of collaboration between scientists and doctors, governments and citizens across the planet. And it would require blood.

Until recently, most blood-serum tests detected antibodies for a single pathogen at a time. But recent breakthroughs have expanded that capability enormously. One example, a method developed at Harvard Medical School in 2015 called VirScan, can detect over 1,000 pathogens, including all of the more than 200 known viruses to infect humans, from a single drop of blood.

[Uncertainty Toolbox](#) is "a python toolbox for predictive uncertainty quantification, calibration, metrics, and visualization", available on GitHub. The toolbox, as per the [accompanying paper](#), was created in order to better calibrate machine learning models. Previous similar projects in this area are: [Ergo](#) and [Python Prediction Scorer](#).

The [Alpha Pundits Challenge proposal](#) was a proposal by the Good Judgement Project to take predictions by pundits, convert their verbal expressions of uncertainty into probabilities, and compare those probabilities to predictions made by superforecasters. Tetlock received some

unrestricted funding from [Open Philanthropy](#) back in 2016, and the grant mentioned the proposal. However, since there isn't more publicly available information about the project, we can guess that it was probably abandoned.

Whenever alpha-pundits balk at making testable claims —like an 80% chance of -2% or worse global deflation in 2016—GJP's ideologically balanced panels of intelligent readers will make good-faith inferences about what the pundits meant. Using all the textual clues available, what is the most plausible interpretation of “serious possibility” of global deflation? GJP will then publish the readers’ estimates and of course invite alpha-pundits to make any corrections if they feel misinterpreted.

GJP Superforecasters will also make predictions on the same issues. And the match will have begun—indeed it has already begun.

For instance, former Treasury Secretary Larry Summers recently published an important essay on global secular stagnation in the Washington Post which included a series of embedded forecasts, such as this prediction about inflation and central bank policies: “The risks tilt heavily toward inflation rates below official targets.” It is a catchy verbal salvo, but just what it means is open to interpretation.

Our panel assigned a range of 70–99% to that forecast, centering on 85%. When asked that same question, the Superforecasters give a probability of 72%. These precise forecasts can now be evaluated against reality.

What we propose is new, even revolutionary, and could with proper support evolve into a systemic check on hyperbolic assertions made by opinion makers in the public sphere. It is rigorous, empirical, repeatable, and backed by the widely-recognized success of the Good Judgment Project based at the University of Pennsylvania

Yearly Housekeeping

I'm trying to improve this newsletter's content and find feedback really valuable. If you could take 2 minutes to [fill out this form](#) and share your thoughts, that would go a long way.

I've moved the newsletter from Mailchimp to [forecasting.substack.com](#), where I've added an optional paid subscription option. Because I conceive of this newsletter as a public good, I'm not planning on offering restricted content, so the main benefit to paid subscribers would be the personal satisfaction of funding a public good.

Note to the future: all links are added automatically to the Internet Archive. In case of link rot, go [here](#) and input the dead link.

Disconfirmed expectancy is a psychological term for what is commonly known as a failed prophecy. According to the American social psychologist Leon Festinger's theory of cognitive dissonance, disconfirmed expectancies create a state of psychological discomfort because the outcome contradicts expectancy. Upon recognizing the falsification of an expected event an individual will experience the competing cognitions, "I believe [X]," and, "I observed [Y]." The individual must either discard the now disconfirmed belief or justify why it has not actually been disconfirmed. As such, disconfirmed expectancy and the factors surrounding the individual's consequent actions have been studied in various settings.

Source: [Disconfirmed expectancy, Wikipedia](#)

2020: Forecasting in Review.

This document contains a series of highlights about forecasting in 2020 which I have gathered after 10 months of writing a forecasting newsletter. I'll write one or two paragraphs for each point, and then list ideas which interested readers can follow up on. As such, this piece can be read either as an accessible superficial summary, as an index of pointers, or as a resource for later years—a snapshot of what was happening in 2020.

Index.

- I. The Phantom COVID-19 Forecasting
- II. Attack of the Metaculus
- III. Revenge of the Crypto Prediction Markets
- IV. Disaster Forecasters—A New Hope
- V. US Election Forecasting Strikes Back
- VI. Return of the Machine Learning
- VII. Platforms Awaken
- VIII. The Last Superforecaster

You can also see this on [substack](#).

I. The Phantom COVID-19 Forecasting

The world in general wasn't prepared to forecast the spread of COVID-19.

"The original IHME model underestimates uncertainty and 45.7% of the predictions (over 1- to 14-step-ahead predictions) made over the period March 24 to March 31 are outside the 95% PIs. In the revised model, for forecasts from April 3 to May 3 the uncertainty bounds are enlarged, and most predictions (74.0%) are within the 95% PIs, which is not surprising given the PIs are in the order of 300 to 2000 daily deaths. Yet, even with this major revision, the claimed nominal coverage of 95% well exceeds the actual coverage. On May 4, the IHME model undergoes another major revision, and the uncertainty is again dramatically reduced with the result that 47.4% of the actual daily deaths fall outside the 95% PIs—well beyond the claimed 5% nominal value." ([Source](#))

Previously, preparedness prediction exercises had been carried out for Ebola. These, however, failed to generalize, because Ebola produces symptoms which are much easier to detect, unlike COVID-19. Irritatingly, media and predictors often confused the detected and the actual cases, as did some forecasting models.

Threads to follow up on:

- [A Case Study in Model Failure? COVID-19 Daily Deaths and ICU Bed Utilisation Predictions in New York State](#)
- [Forecasting s-curves is hard](#)
- [This coronavirus model keeps being wrong. Why are we still listening to it?](#)
- [A Failure, But Not Of Prediction](#)

- [Ioannidis: Forecasting for COVID-19 has failed](#) vs. [Taleb: On Single Point Forecasts for Fat-Tailed Variables](#)
- [How does pandemic forecasting resemble the early days of weather forecasting](#)
- [Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014-15](#)
- [CDC wants states to count ‘probable’ coronavirus cases and deaths, but most aren’t doing it](#)
- [COVID-19 Projections](#)
- [Forecasting the next COVID-19](#)

II. Attack of the Metaculus

Overall Metaculus, a sophisticated forecasting platform and community with a [pretty good track record](#), organized a large number of activities, tournaments and collaborations this past year. Some of the most noteworthy ones follow:

The [AI progress tournament](#), which aims to predict progress in AI, has 50k in rewards. The first round, currently open, contains fairly specific questions which resolve in circa six months.

The [Insight is 20/20](#) was carried out in partnership with the Ben-Gurion University of the Negev. It experimented with rewarding not only the best forecasters, but also those who produced the best arguments and reasoning for their forecasts.

The [Lighting Round Tournament](#) compared forecasts made by infectious disease experts with those made by Metaculus forecasters. Selected results can be found [here](#): Metaculus forecasters generally did better than experts.

On the topic of COVID-19, the [Li Wenliang Forecasting Tournament](#) was organized to forecast the outbreak, and the [Salk Tournament](#) asked for predictions about vaccine research, development and distribution. Metaculus also created a domain dedicated to pandemic predictions: [pandemic.metaculus.com](#), and put together a [dashboard](#) which presented these predictions on the site. See also: [takeaways from Covid forecasting on Metaculus](#) from the winner of the Li Wenliang Forecasting Tournament.

The [El Paso series](#) was launched to help El Paso County, Texas, plan their response to the outbreak, and also rewarded development of tools for scalable forecasting. The first place in the tournament went to Ought, for their development of [Ergo](#); Ought then went on to [partner with Metaculus](#) and to develop [Elicit](#).

Previously, Metaculus had given a [36% probability to a major naturally-originated pandemic by 2026](#), in a question which opened in 2016 and closed on Jan 1, 2020.

Metaculus has also [partnered](#) with The Economist to predict general news in 2021. The [Ragnarök Question Series](#), back from 2018, continues offering probabilities for various catastrophic events. Lastly, [here](#) is a list of forecasts related to Effective Altruism organisations which are currently on Metaculus.

III. Revenge of the Crypto Prediction Markets

Crypto prediction markets bloomed this year, due to the US elections. Getting money into them for the first time generally takes a couple of days, because of the measures taken to verify all users of crypto prediction exchanges to prevent tax evasion.

Because of the initial optimism for decentralized finance applications using blockchain technologies, transaction fees have gone up and down somewhat unpredictably, making

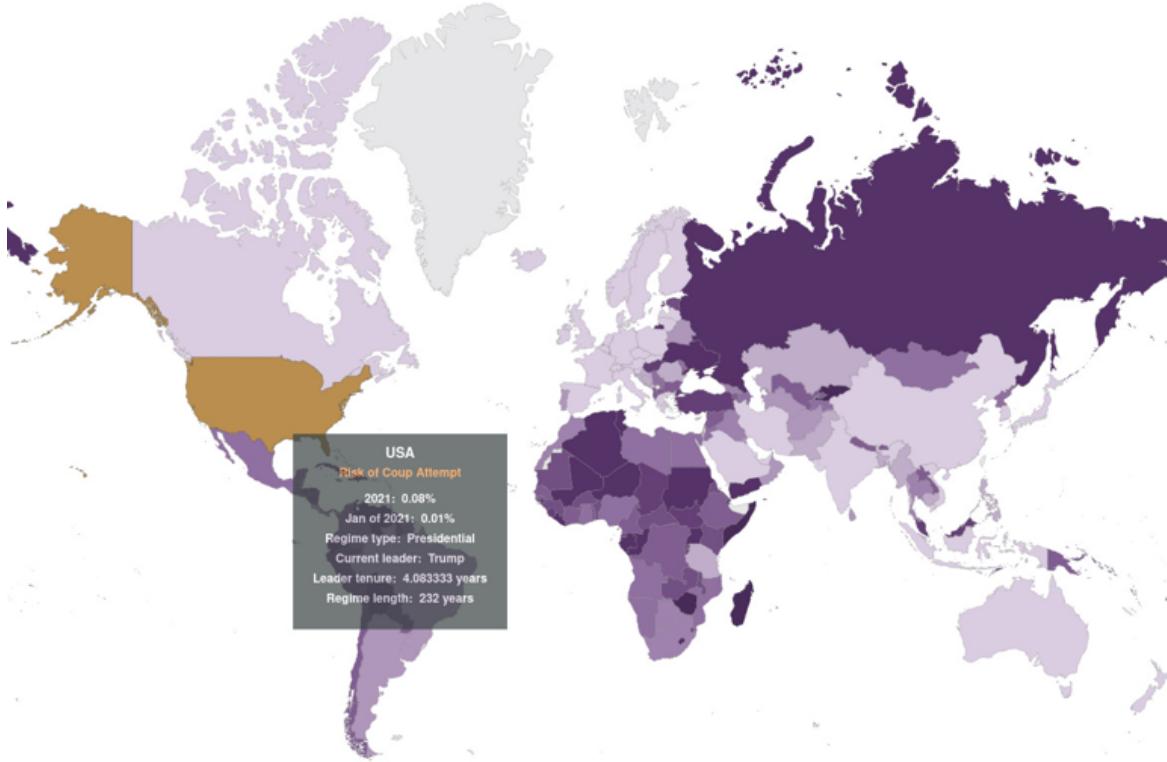
crypto prediction markets such as Augur initially unusable for small players. This led to the development of other crypto prediction markets, such as PolyMarket. They promised to achieve much lower fees by carrying out most transactions in cheaper secondary chains, such as the Matic Network. However, this didn't fix the whole problem because entering and leaving the side-chain still remained expensive. FTX, initially a cryptocurrency derivatives exchange, also offered the possibility to trade US election futures, that is, tokens which could be redeemed for \$0 or for \$1 depending on which presidential candidate won.

One of the most exciting uses of crypto prediction markets was the Omen-based Corona Information Markets, which was designed to produce public and reliable information about COVID-19. However, its markets never saw much volume, and thus their predictions weren't that reliable. Later on, Omen went on to partner with Kleros to deliver decentralized resolutions using a clever scheme where, in essence, judges are incentivized to judge that which the majority of judges will judge (a Keynesian beauty contest).

Threads to follow-up on:

- [Augur](#)
- [Catnip.exchange](#), an interface for Augur.
- [Gnosis](#)
- [Announcing GnosisDAO](#)
- [PolyMarket](#)
- [Polymarket Relayer Community Announcement](#)
- [Matic Network](#)
- [FTX](#)
- [Omen](#)
- [coronainformationmarkets.com](#)
- [Kleros](#)
- [List of prediction markets](#)
- [Prediction Markets' Time Has Come, but They Aren't Ready for It](#)
- [How Accurate Are Prediction Markets?](#)

VII. Disaster Forecasters—A New Hope



The Red Cross and Red Crescent societies have been trying out "forecast-based financing". The idea is to create forecasts and early warning indicators for negative outcomes, such as floods, using weather forecasts, satellite imagery, climate models, etc., and release funds automatically if the forecast reaches a given threshold. This allows the funds to be put to work before the disaster happens in a more automatic, fast and efficient manner.

"In the precious window of time between a forecast and a potential disaster, FbF [Forecast-based Financing] releases resources to take early action. Ultimately, we hope this early action will be more **effective at reducing suffering**, compared to waiting until the disaster happens and then doing only disaster response. For example, in Bangladesh, people who received a forecast-based cash transfer were less malnourished during a flood in 2017." ([Source](#))

Separately, disaster forecasting systems—such as those for floods, hurricanes, famine, locusts, solar flares, volcanoes, etc.—continue to be improved.

Threads to follow-up on:

- [Forecast-based Financing](#)
- [Introductory video](#)
- [Locust-tracking application](#)
- [In Forecasting Hurricane Dorian, Models Fell Short](#)
- [Pan-African Heatwave Health Hazard Forecasting](#)
- [Space Weather Challenge and Forecasting Implications of Rossby Waves](#)
- [USAID's Intelligent Forecasting: A Competition to Model Future Contraceptive Use](#)
- [Forecasting the dividends of conflict prevention from 2020 - 2030](#)
- [How to improve space weather forecasting](#)
- [Coup cast](#)
- [Flood forecasting system didn't help](#)
- [Time to develop a reliable flood forecasting model](#)

- [Google's AI Flood Forecasting Initiative now expanded to all parts of India: Here's how it helps](#)
- [NSF, NASA partner to address space weather research, forecasting](#)
- [Forecasting Solar Flares](#)
- [Forecasting changes in Earth's magnetic field](#)
- [US Congress adopts a plan to consolidate weather catastrophe forecasting](#)
- [Volcano forecasting models](#)

V. US Election Forecasting Strikes Back

US election forecasting was the big story this year, because of the volume of bets it brought. Some threads to follow-up on:

- [Forecasting Newsletter November: United States Presidential Election Post-mortems](#)
- [Forecasting Newsletter December: US Presidential Election Betting](#)
- [Limits of Current US Prediction Markets \(PredictIt Case Study\)](#).
- [Information, incentives, and goals in election forecasts](#)
- [The Primary Model](#)
- [The Long Fork Project](#)
- [Historical Presidential Betting Markets](#)
- ["Biden yes" continues to trade at \\$0.9-\\$0.92 on Polymarket](#) (for an implied probability of 90%, instead of the more reasonable 99%+).

VI. Return of the Machine Learning

Epistemic status: Not an epidemiologist AI researcher.

There were two particularly impressive prediction-like machine learning systems this year, namely GPT-3 and AlphaFold.

I received some feedback that the connection between these advances and forecasting wasn't quite clear, so I thought I'd clarify it. As far as I understand, what GPT-3 does can be thought of as a prediction task: given a prompt, predict its most likely completion. Similarly, AlphaFold is trying to predict the protein structure of a given protein. In contrast, other machine learning systems and training procedures, such as, say, AlphaGo or OpenAI Five, aren't as prediction-like.

We can think of the spectrum from human judgemental forecasting to machine learning as follows:

- Human judgmental forecasting: Complex questions, low amounts of raw data to work with, or the data hasn't been gathered.
 - Example question: "Between 10 July and 31 December 2020, will a firm or paid backup driver operating a self-driving vehicle face criminal charges in relation to an accident involving a self-driving vehicle in the U.S.?"
 - Example method: Do a Fermi estimate with made up numbers.
- Data analysis: Well-posed questions with a medium amount of decently structured data.
 - Example: "Given past sales, what sales volume might this mid-sized company be expected to have in the next year?"
 - Example method: Train a random forest.
- Machine learning: Given vast, vast amounts of structured data, train a machine learning model.
 - Example question: "Given all the products each Amazon user has bought or searched for, output recommendations."
 - Example method: Train a vast machine learning system.

One powerful move for a forecaster is to switch between these levels. For example, given a graph with a very small number of data points, some people would run regressions, but I can sometimes do better by pulling numbers or impressions out of thin air (and, normally, adding more uncertainty). The reverse is also true: Nate Silver could search for structured data to train his US elections model, and beat both pundits and prediction markets. Or, suppose that several million companies each have their own data teams to analyze sales data. You can offer a broader service to all of them, train a superior and more accurate machine learning model on all of their sales data, and perhaps do it more cheaply, because you don't have to pay data analysts.

An interesting development is that machine learning systems are encroaching on data analysis, and data analysis is encroaching on human judgmental forecasting, which is finding new domains. For example, AlphaFold surpasses and augments the biochemists who were previously best at predicting protein shapes, and GPT-3 can sometimes produce reasonable outputs given just a few prompts (though the predictions it produces are so far pretty terrible, I checked). On this note, the nonprofit Ought is trying to skip (or systematize) the process of turning human judgemental forecasting problems into machine learning problems. Although their agenda sounds oddly plausible, it is also very ambitious and they still have a long way to go.

Simultaneously, there have also been some research and forecasting efforts dedicated to predicting AI progress timelines. This has taken various shapes, such as:

- Trying to estimate how many computations per second humans execute, and when machines will reach that point.
- Looking at past technologies to see how often technological progress has been discontinuous, and what the magnitude of past discontinuities has been.
- Eliciting beliefs from informed crowds or experts.
- Incentivizing forecasters to make accurate predictions on the topic of AI progress timelines.
- Checking the accuracy of predictions of technological progress made 10 or more years in advance

Threads to follow up on:

- [Atari, early](#)
- [Language Models are Few-Shot Learners](#)
- [GPT-3](#)
- [AlphaFold](#)
- [OpenAI Five](#)
- [AlphaGo](#)
- [Ought](#)
- [2020 AI Alignment Literature Review and Charity Comparison](#) (search for "#Forecasting")
- [Draft report on AI timelines: Summary](#)
- [Discontinuous progress in history: an update](#) (see also: [A prior for technological discontinuities](#))
- [Forecasting Thread: AI Timelines](#)
- [AI progress tournament](#)
- [Assessing Kurzweil predictions about 2019: the results](#) (see also [Feynman in 1985](#), lucidly answering questions about whether machines will ever be more intelligent than humans.)
- [The Parable of Predict-O-Matic](#)
- [CSET-foretell](#)

IV. Platforms Awaken

CSET-foretell is a new platform which aims to predict geopolitical tensions and transformative technological change and, using the insights gained, attempts to influence policy in the US. CSET (the Center for Security and Emerging Technology), its parent organization, got started with a large grant from Open Philanthropy. I'm back to being the #1 forecaster there, after having momentarily lost the position to user @Hinterhunter. CSET-foretell researchers are gratifyingly superb at metrizing the grand, that is, decomposing and operationalizing big picture questions into smaller ones which can then be forecasted by humans such as myself.

ReplicationMarkets, an effort to study how replication can be best predicted, awarded over \$100,000 in prizes through 2020. They organize prediction markets where traders can buy and sell futures contracts which would pay if a paper replicated, and otherwise wouldn't, and where the best traders are rewarded with real money. In addition, they also organized survey rounds in which forecasters predicted alone, and which paid out sooner. Two other, smaller, similar efforts in this area are The Pipeline Project, and the Social Science Prediction Platform. As for the second, as far as I understand, although it is open to anyone, only graduate students can receive their symbolic \$25 reward.

Otherwise, Facebook announced Forecast, a forecasting app. Twitch also added prediction functionalities to their platform.

Threads to follow-up on:

- [CSET-foretell](#). Their decomposition of high-level questions into many specific sub questions is superb; more can be read about their approach here: [Future Indices: How Crowd Forecasting Can Inform the Big Picture](#) or their [blog](#).
- [ReplicationMarkets](#)
- ReplicationMarkets: [Prediction Market for COVID-19 Preprints](#)
- [Are replication rates the same across academic fields? Community forecasts from the DARPA SCORE programme](#)
- [Social Science Prediction Platform](#)
- [Replication Markets: Can You Predict Which Social Science Papers Will Replicate?](#)
- [Estimating the deep replicability of scientific findings using human and artificial intelligence](#)
- [What's Wrong with Social Science and How to Fix It: Reflections After Reading 2578 Papers](#).
- [The Pipeline Project](#)
- [Facebook's Forecast](#)
- [Forecast Update: Making Forecast Publicly Available in the US and Canada](#)
- [Azul Forecasting](#)

VIII. The Last Superforecaster

Good Judgement Inc. is the organization which grew out of Tetlock's research on forecasting, and out of the Good Judgement Project, which won the IARPA ACE forecasting competition, and brought about the research covered in the book *Superforecasting*. Good Judgement Inc. also organizes Good Judgement Open, a forecasting platform with a focus on serious geopolitical questions which is used to identify Superforecasters, whose predictions are then offered as a proprietary service.

This year, Good Judgment 2.0—a new research project by Tetlock et al.—participated in IARPA's FOCUS (Forecasting Counterfactuals in Uncontrolled Settings) tournament, and to do this, started a new R&D project, Good Judgment 2.0. Otherwise, there just isn't much high-quality publicly available information about what they've been up to recently.

Meanwhile, while the CultivateLabs platform—used by Good Judgment Open and CSET-foretell—has remained relatively static, other forecasting platforms and projects have been making their own advances. Metaculus and Foretold have continuous questions in which

forecasters produce probability distributions over a range, rather than probabilities over binary outcomes. Ozzie Gooen, of Foretold and Guesstimate fame, has also been experimenting with forecasting using functions that return probability distributions—rather than forecasting probability distributions directly. Elicit's and Metaculus' predictions can be embedded on other webpages. On the prediction markets front, Polymarket and Augur allow users to trade on their implied probability distributions using scalar markets. Augur further allows users an option to trade on an "ambiguous" resolution as one of the explicit outcomes. And Gnosis has been experimenting with futarchy: using prediction markets to make decisions.

Threads to follow up on:

- [The One Good Overview](#) of what Good Judgment (2.0) has been up to. Thanks to Walter Frick & Quartz for relaxing the paywall so that access only requires an email. Warning: long.
- [Tetlock's Twitter](#)
- [Good Judgement Open](#)
- [Good Judgement Inc.](#)
- [IARPA's Aggregative Contingent Estimation \(ACE\) Program](#)
- [IARPA's FOCUS tournament](#)
- [Good Judgement 2.0 project](#)
- [Good Judgement Dashboard](#)
- [Multivariate estimation & the Squiggly language](#)
- [Elicit](#)
- [Augur scalar markets](#)
- [GnosisDAO](#)
- [Futarchy](#)
- [A review of Tetlock's 'Superforecasting' \(2015\)](#), by Dominic Cummings

Forecasting Newsletter: January 2021

Highlights

1. Veteran PredictIt trader writes a [pretty good guide](#) on how to make money on prediction markets.
2. [Metaculus](#) and [Hypermind](#) both have new COVID-19 forecasting tournaments.
3. I created a [search engine for probabilities](#).

Index

- Highlights
- Prediction Markets & Forecasting Platforms
- In The News
- Long Content
- Hard To Categorize

Sign up [here](#) or browse past newsletters [here](#).

Prediction Markets & Forecasting Platforms

Hypermind is an American-French forecasting platform with a somewhat outdated and clunky interface. They have a new [COVID-19 Recovery](#) contest with \$7000 in promised prizes so far, with the amount set to increase as more questions get added. The contest is sponsored by the Open Philanthropy Project. Hypermind is somewhat difficult to navigate, so you might only be able to find the contest if you create an account and look around.

Metaculus has a new [COVID-19 Forecasting](#) contest. From the description:

"The goal of this project is to provide probabilistic predictions of the U.S. COVID-19 outbreak to support public health decision making at the federal and state level.

At the end of each month we will share a summary report with the Council of State and Territorial Epidemiologists, members of the Centers for Disease Control and Prevention, all members of MIDAS (Modeling of Infectious Disease Agent Study), and make this report available for public consumption."

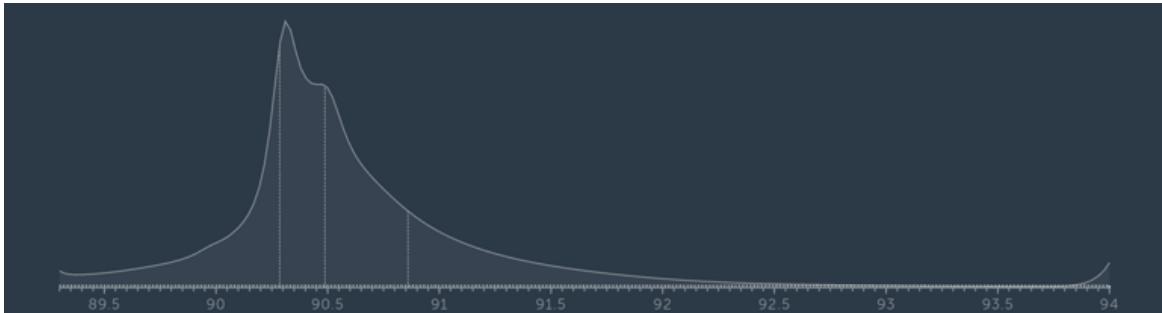
Metaculus also published an open letter on the [urgent need for expanded surveillance and forecasting of novel SAR-CoV-2 variants](#). An opinion piece saying the same thing is also available on [The Hill](#):

"Efforts are already being made to characterize and understand the infectivity properties and immunological consequences of these new variants. However, as was the case at the start of the pandemic, most countries remain extraordinarily uncertain as to (1) the extent to which these novel variants are spreading and (2) the likelihood as to whether and when these new variants will become predominant.

Unfortunately, these issues do not appear as if they will be extensively addressed in the immediate future. The U.S. CDC, for instance, is currently only planning on having each state send it 'at least 10 samples' on a biweekly basis for sequencing and further characterization. This is woefully inadequate genomic surveillance — we are in the dark.

We are calling for a massive increase in genomic sequencing, monitoring, data sharing, and probabilistic forecasting so we can have a detailed understanding of where these new variants are circulating and how rapidly they are increasing as a proportion of all cases."

On the negative side, Metaculus's current editor could use some improvement. For example, consider the following aggregate prediction on the [state of the art performance on the SuperGLUE AI benchmark](#):



The current state of the art performance is [90.3](#) in the SuperGLUE benchmark, making it extremely unlikely that the end result will fall below that number. But, the Metaculus's aggregate prediction gives a 25% chance to the state of the art falling below that number at question resolution time. This is because the Metaculus interface makes it annoying, or directly impossible, to create one-sided tails.

[Omen](#) announced an integration with [API3](#). This will allow for obtaining generally superior resolutions for the price of almost any cryptocurrency. However, Omen is seeing very low trade volumes and very low numbers of active questions.

In contrast, [Polymarket](#) has been doing quite well with regards to trade volume. They resolved some of their presidential succession questions, and have probably managed to keep some of the new users in the aftermath. For a while, Polymarket was "an unlimited passive income stream for people who still have their frontal lobes intact" ([source](#)). For example, [Will Joe Biden be inaugurated as President of the USA on January 20th, 2021?](#) traded at 85 to 93% (!).

But now—unlike in the previous edition of this newsletter—I don't think that there are any markets which are egregiously wrong after taking into account fees and the hassle of moving relatively small amounts of money into Polymarket. That said, the "No" position on [Will Donald Trump be President of the USA on March 31, 2021?](#) is still trading at 97 to 98% (after fees). And, a 2 to 3% return per month, particularly if compounded for a year, still looks pretty good.

[Augur](#), a more decentralized cryptocurrency-based prediction market, has also successfully resolved various US election questions, and has also done better in terms of volume, particularly since its new interface, [catnip.exchange](#), sprung up. However, I haven't been following them closely.

In other news, I created a [search engine for probabilities](#). It currently aggregates forecasts from PredictIt, Polymarket, Omen, Metaculus, Good Judgment Open, CSET-foretell, Elicit, PredictionBook (through Elicit) and Hypermind. You can access a demo [here](#), or browse a GitHub repository and find out the location of selected API endpoints [here](#). To get a feel of how it works, I suggest searching for "Trump", "China", or "semiconductors". Tentatively, I'll keep both the search engine and the json/csv endpoints updated once a day for the next month. I consider this to be in very early beta: comments and suggestions are welcome.

Trump

Found 15 records matching in Metaforecasts				
URL	PLATFORM	BINARY Q...	PERCENTAGE	# FORECASTS
https://polymarket....	PolyMarket	✓	5.4501%	265.00
https://www.peric...	PredictIt	✓	2%	
https://www.metac...	Metaculus	✓	33%	63
https://www.metac...	Metaculus	✓	16%	1116
https://elicit.org/bi...	Elicit	✓	10.55%	38
https://www.metac...	Metaculus	✓	3%	244
https://polymarket....	PolyMarket	✓	4.9602%	1144.00
https://polymarket....	PolyMarket	✓	5.0066%	553.00

In the News

[Forecasting the New Administration's Impact on Defense](#). Despite being quite badly formatted, this piece by a former vice president of combat avionics at Northrop Grumman provides deep expertise and insight on the future shape of US defense spending under the Biden administration. The piece doesn't provide explicit probabilities, but it does give a sense of which scenarios are most likely and which are most worth paying attention to.

[Vox looks back at their forecasts from 2020](#), and they compare favorably to [Metaculus](#)'s ([source](#)). Vox also [offers new predictions for 2021](#).

[Radar technology that could revolutionize hurricane forecasts hits major setback](#). The US's National Science Foundation considered the price tag of \$70 million insufficiently justified.

"An airborne phased-array radar system consists of thousands of transmitters and receivers spread across four square arrays strategically placed on an aircraft's fuselage. They scan the sky and "can provide unprecedented detailed observations of the dynamics and microphysics of high-impact storms," according to an NCAR fact sheet. The data collected by the phased-array radar, when integrated into computer models, could improve forecasts for hurricanes and other hazards investigated by aircraft, including non-hurricane severe weather and winter storms."

[Airports explore new ways to forecast travel amid the pandemic](#) by looking at new indicators, such as the number of people who search for the opening times of the Statue of Liberty, or for rental cars.

[AI Startup Sees Opportunity Forecasting Pandemic-Era Consumer Demand](#) using proxies which other companies don't yet use as much, such as internet searches.

[Betting Against QAnon](#) proved particularly profitable for some of the PredictIt traders.

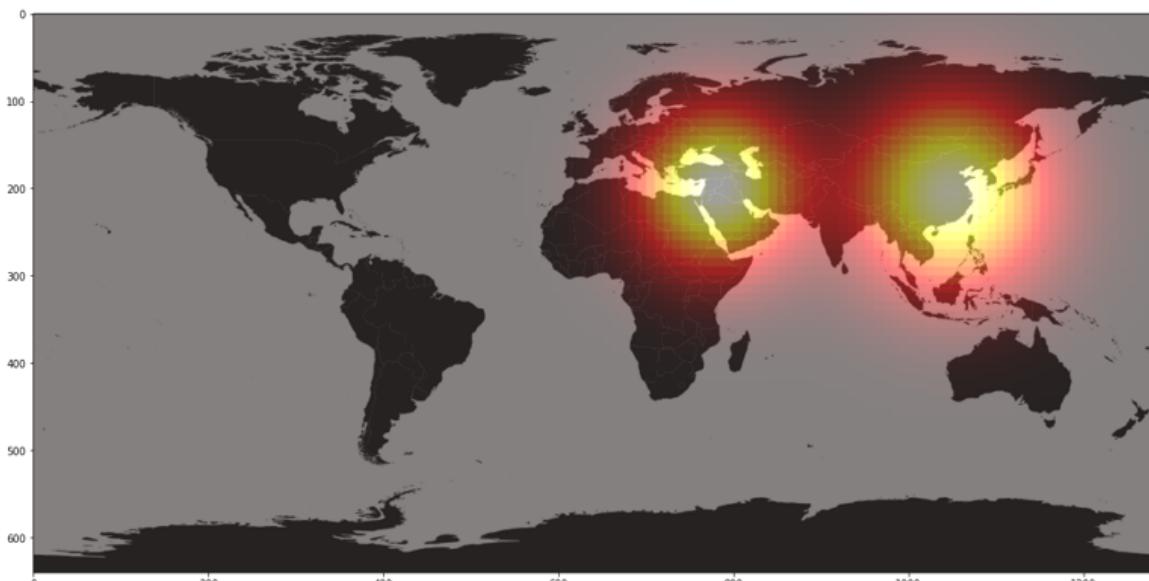
[Superforecasters have a look at the end of Covid](#) in Britain.

Hard to Categorize

A small US city [deliberates](#) about paying for an expansion to a [gunfire locator](#) which would not only detect and report shots fired, but also direct police officers to areas where incidents are predicted to be likely to happen. See also: [Minority Report](#).

[Rootclaim](#) is a site which comes up with Bayesian calculations for public interest questions. For example, here is their page on [the source of COVID-19](#): they start with a reasonable prior and then legibly update their initial prediction with each piece of evidence they consider. That said, their conclusion differs from that of [Metaculus](#) and from that of casual [discussion](#) between several superforecasters on Twitter.

Metaculus user [Ege Erdil](#) has produced a heatmap of predicted locations for World War 3 putting together the results of two questions: [If there is a WW3, what latitude will it start in?](#) and [If there's a WW3, what longitude will it start in?](#). The source code used to produce the image below is available [here](#). Because the latitude and longitude are given as separate variables, the code uses some kernel wizardry to try to find their degree of correlation, which might introduce some mistakes.



WW3 according to Metaculus.

Another Metaculus user and top 50 forecaster, [SimonM](#), has created a page, [Metaculus Extras](#), which presents various statistics about the platform, such as a list of top comments, an h-index (!), and a [timeline](#) of Metaculus community predictions.

Long Content

[A new paper \(summary\)](#) tries to quantify by how much entrepreneurs are overconfident when presenting forecasts to potential investors. The authors found ~15% overconfidence in founder CEOs, and ~27% for non-founder CEOs.

[Forecasting the future risk of dengue epidemics facing climate change in New Caledonia, South Pacific.](#)

"Over the last decade, the toll of dengue fever has increased in New Caledonia, raising questions about the future of the disease in this French island territory located in the South Pacific. Climate has a strong influence on dengue through its influence on the ecology of the vector and the viral cycle. Several studies have explored the link between climate and dengue in New Caledonia, with the aim of explaining and predicting dengue outbreaks. None of these studies have explored the possible outcome climate change will have on the risk of dengue fever in New Caledonia. This is the goal of this study, through projections of rainfall and temperature and the selection of an appropriate prediction target for our statistical model, we assess the climate-induced risk of dengue outbreaks up to the 2100 horizon. We prove that the inter-annual risk of dengue outbreaks in New Caledonia will raise, according to all the greenhouse gas emission scenarios and according to the high emission scenario, dengue fever will become an endemic disease in New Caledonia."

A [recent working paper](#) by the Federal Reserve Bank of Philadelphia introduces a class of disagreement measures for probability distribution forecasts based on the [Wasserstein metric](#) (also known as the [Earth mover's distance](#)).

[Against essential and accidental complexity](#).

"In the classic 1986 essay, No Silver Bullet, Fred Brooks argued that there is, in some sense, not that much that can be done to improve programmer productivity. His line of reasoning is that programming tasks contain a core of essential/conceptual complexity that's fundamentally not amenable to attack by any potential advances in technology (such as languages or tooling). He then uses an Ahmdahl's law argument, saying that because $1/X$ of complexity is essential, it's impossible to ever get more than a factor of X improvement via technological improvements. Towards the end of the essay, Brooks claims that at least $1/2$ (most) of complexity in programming is essential, bounding the potential improvement remaining for all technological programming innovations combined to, at most, a factor of 2. Let's see how this essential complexity claim holds(...)"

[Mining the silver lining of the Trump presidency](#): A retrospective look at the Trump tweets' markets in PredictIt.

"For the better part of the last four politically insane years, a community of gamblers wagered stupid amounts of money betting on a simple question: How many times would Donald Trump tweet this week? The game ended for us before it ended for the President, but now that it's completely over, I feel this tiny corner of Internet weirdness deserves some remembrance. After all, there are very few other people on this planet that understood Trump's twitter habits – and by extension, Trump himself – more than the people who bet on them."

That same author has what seems to be a [pretty good guide](#) on how to bet on prediction markets if one is optimizing for making money. Something I personally was doing wrong was holding until the end and considering prices in isolation, that is, asking myself "is this price wrong?" rather than "is this price the most wrong it will be?"

The UK's [National Risk Register](#) is a document which "provides information on the most significant risks that could occur in the next two years and which could have a wide range of impacts on the UK." Besides this, it also contains a pretty good categorization scheme for risks.

Metaculus lore tells of a [legendary comment](#) by user [@travisfisher](#). Written on Jan 24, 2020, under [Will the world population increase every year for the next decade?](#), it reads:

The Wuhan Coronavirus is looking like a pandemic event that could be serious enough to threaten this outcome.

Note to the future: All links are added automatically to the Internet Archive. In case of link rot, go [there](#) and input the dead link.

Probability should not be introduced through games of chance. These games are artificial, & give the impression that probability is mostly objective & irreducible (aleatoric). The real problems we face almost always require probability that is subjective & reducible (epistemic).

Source: [@maosbot](#)

Forecasting Newsletter: February 2021

Highlights

- Biden orders the creation of a [National Center for Epidemic Forecasting and Outbreak Analytics](#)
- Americans get a [clone](#) of Betfair/Smarkets, expected to be marginally better than PredictIt
- Hypermind has a new forecasting tournament on the [future of AI in 2023](#)

Index

- Prediction Markets & Forecasting Platforms
- In The News
- Recent Blog Posts
- Personal Forecasts
- Hard To Categorize
- Long Content

Sign up [here](#) or browse past newsletters [here](#).

Prediction Markets & Forecasting Platforms

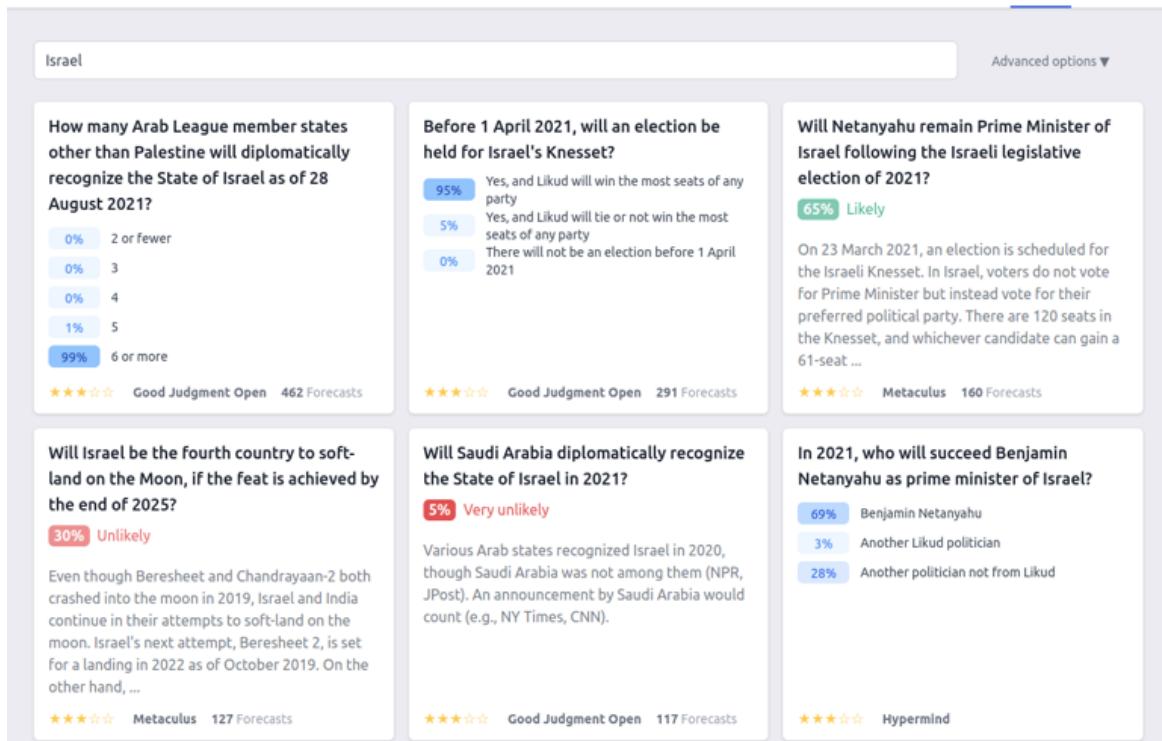
A new US company, [Kalshi](#), has gotten [regulatory approval](#) from the Commodity Futures Trading Commission to create a betting platform. I don't really have many thoughts given that they haven't launched yet. I expect them to use newer technologies than PredictIt just because they are newer, and I expect them to have somewhat lower fees because it would make business sense. They are planning to open in March. From the [Wall Street Journal](#) (unpaywalled [archive link](#)):

Alfred Lin, a partner at Sequoia and a Kalshi board member, said Kalshi's embrace of regulation was one of the reasons his firm invested in the startup. "They're taking regulation fairly seriously," he said. "Companies that move fast and break things are not going to work in this regulated environment."

[Hypermind](#) has a new forecasting tournament: [Where will AI be in 2023?](#), with a prize pot worth \$7,000 so far.

Forecasters on Good Judgment Open have the opportunity to receive feedback from superforecasters if they participate in the [Think Again challenge](#), make 10 predictions and complete a survey.

With [QURI](#), I've been improving [Metaforecast](#), a search tool for probabilities. It now has more prediction platforms, a nicer interface, and more search options. Readers might be interested in [COVID predictions](#), readers from the EA movement might be particularly interested in [all GiveWell and OpenPhilanthropy predictions](#).



[Metaforecast search for "Israel!"](#)

In the News

Biden ordered the creation of a [National Center for Epidemic Forecasting and Outbreak Analytics \(secondary source\)](#). The new agency looks somewhat related to a previous proposal mentioned in this newsletter: [Forecasting the next COVID-19](#).

[Suboptimal demand forecasting](#) for semiconductor chips has led to pausing automobile production in the US. On the one hand, automakers struggle to compete for chips against more profitable tech products—e.g., iPhones—on the other hand, US sanctions on China’s Huawei and [SMIC](#) put even more pressure on semiconductor production capacity.

[The European Central Bank will be holding its 11th Conference on Forecasting Techniques as an online event on 15 and 16 June 2021.](#)

Hewlett Packard has built a new [supercomputer](#) dedicated to weather forecasting for the U.S. Air Force. The new system advertises a peak performance of 7.2 petaflops. This is comparable to estimates of the [human brain](#), and around two orders of magnitude lower than the [fastest supercomputer](#).

[Future Returns: Using the Past to Forecast the Future of the Markets](#). An analyst at [Fidelity](#) looks at the historical base rate for market behavior in situations similar to the current, COVID-19-affected, performance.

[Where The Latest COVID-19 Models Think We're Headed — And Why They Disagree](#), by FiveThirtyEight

Recent blog posts

[Boring is back, baby!](#) Experienced PredictIt bettor discusses the future profitability of political predictions:

The political betting community has been quietly dreading the potential boringness of the Biden presidency – without politics being so crazy, engagement should fall off and so should the deposits of new accounts coming in to bet on whatever wild stuff Trump was up to next

I'd more or less written off this year as one in which I'd be happy to earn a third what I did last year on PredictIt and maybe try doing some grown-up work or something (lol, as if). Then it turns out January was one of the most interesting months in politics of the entire Trump presidency (to put it mildly) and engagement has remained fairly substantial. But that doesn't mean the doldrums aren't coming.

The following three articles, among others, won a “[Forecasting Innovations Prize](#)” I had previously co-organized under QURI.

[Crowd-Forecasting Covid-19](#) describes the results of a COVID-19 crowd-forecasting project created during the author's PhD. This is probably the one app in which human forecasters can conveniently forecast different points in a time series, with confidence intervals. The project's forecasts were submitted to the German and Polish Forecast Hub, and they did surprisingly well in comparison with other groups. They are [looking for forecasters](#), and will soon expand to cover 32 European countries as part of the yet-to-be-launched [European Forecast Hub](#).

[Incentivizing forecasting via social media](#) explores the implications of integrating forecasting functionality with social media platforms. The authors consider several important potential issues at length, propose possible solutions, as well as give recommendations regarding next steps. The scenario they consider—if it were to occur—could possibly have a large impact on the “information economy”.

[Central Limit Theorem investigation](#) visualizes how quickly the central limit theorem works in practice, i.e., how many distributions of different types one has to sum (or convolve) to approximate a Gaussian distribution in practice (rather than in the limit). The visualizations are excellent and give the readers intuitions about how long the central limit theorem takes to apply. As a caveat, the post requires understanding that the density of the sum of two independent variables is the convolution of their densities. That is, that when the post mentions “the number of convolutions you need to look Gaussian”, this is equivalent to “the number of times you need to sum independent instances of a distribution in order for the result to look Gaussian”. This point is mentioned in an [earlier post](#) of the [same sequence](#).

I stumbled upon [Alert Foxes](#), a blog with a few forecasting posts by Alex Foster (perfect anagram!). I particularly enjoyed the decompositions of his predictions on [US election questions](#).

Vitalik Buterin writes about his experience [betting on the US election](#) using crypto prediction markets.

[AstralCodexTen](#)—previously [SlateStarCodex](#), a blog I hold in high regard and which is probably known by everyone on LW—has started a weekly series discussing forecasting questions ([1](#), [2](#), [3](#)).

Personal forecasts

A piece of [feedback](#) I got at the end of last year about this newsletter was to talk more about my own predictions, so here are two which I recently got wrong and one that I got right:

The first one was [Will Kim Kardashian or Kanye West file for divorce before March 1, 2021?](#). After some investigation, I thought that they would try to time the divorce to maximize news about the last season of [Keeping Up with the Kardashians](#), and was quite surprised when they didn't. Other bettors were also surprised, as the price on Polymarket looked as follows:



PolyMarket prices for question "Will Kim Kardashian or Kanye West file for divorce before March 1, 2021?"

The second prediction I got grievously wrong was [How much new funding will facial recognition companies raise between July 1 and December 31, 2020, inclusive?](#), on CSET-foretell. With the passage of time I updated away from the option "Less than \$200 million", which ended up being chosen for resolution. The resolution source, Crunchbase, describes [Acceso Digital](#) as "a developer of facial recognition and identification technology created to solve document and process management", but doesn't classify it in the "facial recognition" category. In September 2020, Acceso Digital raised R\$580M (circa \$90 million), which would have been enough to raise the final question resolution to the next category ("More than \$200 million but less than or equal to \$500 million").

Thirdly, I assigned a 50% probability to winning an EA forum prize for a [research project](#), which I [did](#).

Hard to Categorize

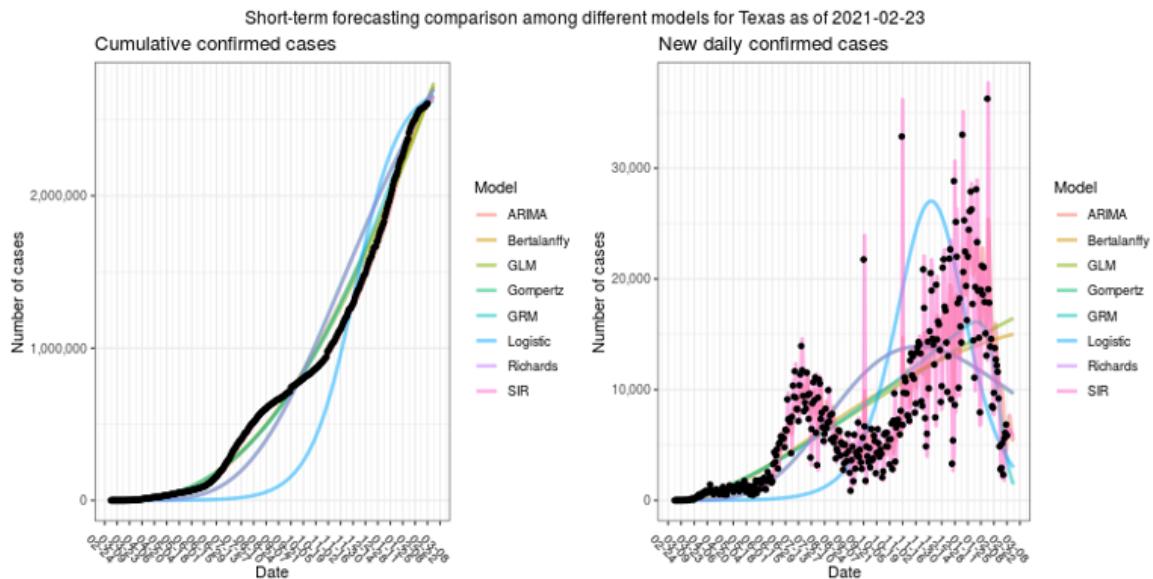
The [Illinois Commission on Government Forecasting and Accountability](#) is a government agency in charge of making e.g., revenue predictions. Judging by its webpage, it seems somewhat outdated. A similar agency in [California](#) appears to be more up to date. It might be interesting for platforms like Metaculus to try to partner with them.

[Meta and consensus forecast of COVID-19 targets](#) ([secondary source](#)) provides a variety of forecasts about COVID. They provided forecasts about US deaths conditional on vaccination rates, which could have been particularly action-guiding. They also find that forecasts which aggregate predictions from infectious disease experts and “trained forecasters” have wider uncertainty intervals than the [COVID-19 Forecast Hub](#).

[Upstart](#), a company which uses machine learning/data analysis to predict loan repayment, is looking for one or more forecasters with a good track record to do some consulting work. If you're interested, let me [know](#).

Long Content

[Evaluating Short-term Forecast among Different Epidemiological Models under a Bayesian Framework](#) ([supporting data](#), [webpage](#)). The authors notice that the relative merits of different epidemic forecasting methods and approaches are difficult to compare. This is because they don't normally have access to the same data or computational capacity in the wild. The authors set out to carry out that comparison themselves, but they don't arrive to any sharp conclusions, other than [ARIMA](#) not being able to keep up with stochastic approaches.



from 'Evaluating Short-term Forecast among Different Epidemiological Models under a Bayesian Framework'

We calibrate stochastic variants of growth models and the standard SIR model into one Bayesian framework to evaluate their short-term forecasts.

Broadly speaking, there are five types of approaches to forecasting the number of new cases or the expected total mortality caused by the COVID-19: 1) time-series forecasting such as autoregressive integrated moving average (ARIMA) [...]; 2) growth curve fitting based on the generalized Richards curve (GRC) or its special cases [...]; 3) compartmental modeling based on the susceptible-infectious-removed (SIR) models or its derivations [...]; 4) agent-based modeling [...]; 5) artificial intelligence (AI)-inspired modeling.

There has been a growing debate amongst researchers over model performance evaluation and finding the best model appropriate for a certain feature (cases, deaths, etc.), a particular regional level (county, state, country, etc.), and more. Fair evaluation

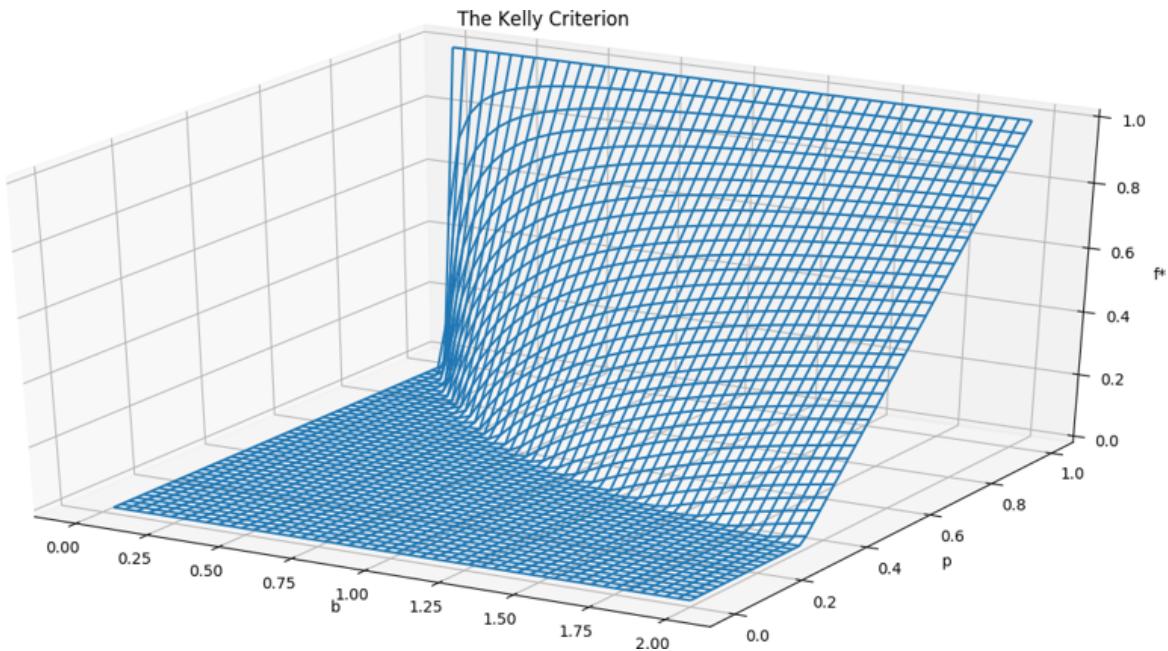
and comparison of the output of different forecasting methods have remained an open question, since models vary in their complexity in terms of the number of variables and parameters that characterize the dynamic states of the system.

Although a comparison of predictive models for infectious diseases has been discussed in the literature, to our best knowledge, no existing work systematically compares their performances, particularly with the same amount of data information.

None of the models proved to be golden standards across all the regions in their entirety, while the ARIMA model underperformed all stochastic models proposed in the paper

Comparing weather forecasts in Tasmania now to those made 30 years ago, a news article mentions that the amount of available data has increased [13.5 million times](#).

[The Kelly Criterion, visualized in 3D](#) .



The Kelly Criterion in 3D, visualized by lsusr. Source: Less Wrong

On the topic of Kelly, see also [Kelly isn't \(just\) about logarithmic utility](#), [Kelly *is* \(just\) about logarithmic utility](#) and [Never Go Full Kelly](#).

[The EpiBench Platform to Propel AI/ML-based Epidemic Forecasting: A Prototype Demonstration Reaching Human Expert-level Performance \(secondary source\)](#).

During the COVID-19 pandemic, a significant effort has gone into developing ML-driven epidemic forecasting techniques. However, benchmarks do not exist to claim if a new AI/ML technique is better than the existing ones. The "covid-forecast-hub" is a collection of more than 30 teams, including us, that submit their forecasts weekly to the CDC.

It is not possible to declare whether one method is better than the other using those forecasts because each team's submission may correspond to different techniques over the period and involve human interventions as the teams are continuously changing/tuning their approach. Such forecasts may be considered "human-expert" forecasts and do not qualify as AI/ML approaches, although they can be used as an indicator of human expert performance.

We are interested in supporting AI/ML research in epidemic forecasting which can lead to scalable forecasting without human intervention. Which modeling technique, learning strategy, and data pre-processing technique work well for epidemic forecasting is still an open problem. To help advance the state-of-the-art AI/ML applied to epidemiology, a benchmark with a collection of performance points is needed and the current "state-of-the-art" techniques need to be identified. We propose EpiBench a platform consisting of community-driven benchmarks for AI/ML applied to epidemic forecasting to standardize the challenge with a uniform evaluation protocol.

In this paper, we introduce a prototype of EpiBench which is currently running and accepting submissions for the task of forecasting COVID-19 cases and deaths in the US states and We demonstrate that we can utilize the prototype to develop an ensemble relying on fully automated epidemic forecasts (no human intervention) that reaches human-expert level ensemble currently being used by the CDC.

In an experiment, the researchers compared 3 AI and machine learning forecasting methods and 30 methodologies pulled from published research using EpiBench. They found that while many of the forecasts reportedly used the same model (SEIR), they predicted "drastically" different outcomes. Moreover, two methodologies identical except that one smoothed data over 14 days versus the other's 7 days varied "significantly" in their performance, suggesting that data preprocessing played a nontrivial role.

Note to the future: All links are added automatically to the Internet Archive. In case of link rot, go [here](#) and input the dead link.

"I never think of the future. It comes soon enough".

Albert Einstein, said probably as a joke ([source](#)).

Forecasting Newsletter: March 2021

Highlights

- OpenPhilanthropy releases a report on [outside view perspectives on the likelihood of AGI](#).
- Jason Matheny, previous director of IARPA, CSET, is now a ¿senior? [official in the Biden administration](#).
- Astral Codex Ten considers [Trapped Priors As A Basic Problem Of Rationality](#).

Index

- Prediction Markets & Forecasting Platforms
- In The News
- Recent Blog Posts
- Hard to Categorize
- Long Content

Sign up [here](#) or browse past newsletters [here](#).

Prediction Markets & Forecasting Platforms

[Numerai](#) is a distributed, blockchain-based hedge fund. Users can either [predict on free, but obfuscated data](#), or [use their own data and predict on real world companies](#). After the users stake cryptocurrency on their predictions, Numerai buys or sells stocks in proportion to each prediction's stake and then stake cryptocurrency on their predictions. The fund observes how well the predictions do. Then it increases the stake of those who did well and burns part of the stake of those who performed badly. Numerai's users currently have around \$12.5 million staked.

CSET's Founding Director [Jason Matheny](#) is now a ¿senior? [official in the Biden administration](#). In his past life, he did some pioneering work on cultured meat, then was a Program Manager of IARPA's [Aggregative Contingent Estimation \(ACE\) program](#) (of Good Judgment fame), before becoming director of IARPA. In recent times, he founded the [Center for Security and Emerging Technologies](#) (CSET.)

CSET Foretell is launching a [Pro Forecaster Program](#) in April 2021, which means it will start paying its forecasters. They are offering to pay \$200/month (each) to 50 selected forecasters. The total payout, which comes to \$120k yearly, competes with Replication Markets as one of the largest forecaster reward budgets.

Pro Forecasters will be paid to make forecasts that contribute to our research and analysis for policymakers. Invitations have been sent to current Foretell users, and we are now accepting applications for the remaining spots. *Anyone who can show a proven track record as a top forecaster on Good Judgment Open, Metaculus, or a similar crowd forecasting site may apply* (emphasis mine).

Using early [data](#), CSET Foretell [finds](#) that its crowd outperforms historical projections.

Personally, I ended up on place #5 out of 646 on the [first season's leaderboard](#), and my team, Samotsvety Forecasting, comprised out of Eli Lifland, Misha Yagudin and myself, completely outpaced all other teams:

Rank	Team	Scored Questions	Overall Score
1.	Samotsvety Forecasting	24	-0.994
2.	Shaun's Team	26	-0.46
3.	Roger's Data Science Team	2	-0.04
4.	Foster Tech Club	2	-0.016
5.	Georgetown IRC	1	-0.007
6.	Wake Forest University	2	0.014
7.	CSET	30	0.104
8.	Foretellers	26	0.118
9.	Drew's Team	14	0.178
10.	Sean's Team	24	0.219
11.	Mauricio's Team	21	0.288
12.	Ariana's Team	22	0.912
13.	Yifan's Team	16	1.26
14.	Molly's Team	24	1.954
15.	Shalin's Team	16	1.997

CSET-Foretell: [Team leaderboard](#) at the end of the first season.

[Omen v2 launches](#). Crucially, they are moving to a subchain. Trades will be cheaper, once they are made inside the subchain, yet this comes at a cost—the process of moving currency to a subchain is cumbersome. They have added more questions, but the platform still remains small.

[Here](#) is a short explanation of how Catnip works, and what the differences between Catnip and Augur are.

Hypermind has a [new forecasting tournament](#)—on the state of AI in 2030—with relatively low rewards of \$3000.

I've continued to improve [Metaforecast](#):

- I added MichaelA's excellent [Database of Existential Risk Estimates](#).
- I also added [Ladbrokes](#), [WilliamHill](#), [Estimize](#) and [FantasySCOTUS](#), “the leading Supreme Court Fantasy League,” which covers cases the US Supreme Court is expected to address this year.
- I marginally improved the search, h/t Peter Hartree for a detailed [pull request](#).
- Ozzie Gooen and I wrote [Introducing Metaforecast: A Forecast Aggregator and Search Tool](#) (also on [LessWrong](#)).
- Metaforecast also saw [some activity](#) on [twitter](#).

While [sniffing](#) the requests from WilliamHill and Ladbrokes, I found out about [OpenBet](#). OpenBet is a provider of software infrastructure for many major European betting houses: WilliamHill, Betfair, PaddyPower, Ladbrokes, etc. They package that software infrastructure with extremely [addictive games](#). OpenBet has a [“Corporate Social Responsibility”](#) page and

various responsible gaming [accreditations](#). But despite these accreditations, the platform's business model appears to rely on making the bettors become addicted: As explained on the —sinisterly named —“[omni-channel](#)” page “multi-channel customers have proven to be 38% more profitable than single channel customers”. It’s unclear to me whether the fact that they manage the infrastructure for so many European betting houses is problematic with regards to the EU’s [antitrust policy](#); I’d give it a 10 to 30% chance.

That modus operandi stands in contrast to current cryptocurrency-based prediction markets, which have much cleaner interfaces and don’t appear to indulge in predatory profit-chasing, e.g., compare [Omen’s frontend](#) with that of [Betfair](#).

In a sad turn of affairs, [Polkamarkets](#), a new prediction market still under development, might also be aiming to capture profit from the addictive gamification aspect of betting within the crypto-prediction market ecosystem. Nonetheless, Polkamarkets hasn’t launched yet and, on the positive side, it promises higher frequency markets and faster resolution times, so it’s still too soon to judge whether it will be a net positive project.

Metaculus

Metaculus is [hiring](#).

There has been some [discussion about, and criticism of](#), the Metaculus scoring rule. A Metaculus co-founder answers with [A Primer on the Metaculus Scoring Rule](#).

[MetaculusExtras](#) has added new features, like the daily and cumulative [number of predictions](#) on the site or the [number of points per question](#) per user. [SimonM](#), of MetaculusExtras fame, also [extracts](#) the top comments (i.e. the most upvoted and slightly curated) made in March on Metaculus:

- misha lays out the [various different odds for the Olympics](#) on different platforms. Relatedly, and not on Metaculus, see [this twitter thread](#) by Brian Lui about a trader on FTX who is taking all bets.
- zc points out some examples of countries having [fantastic longevity improvements](#)
- ege_erdil calculates the [base rate for resignations for politicians resigning after accusations of misconduct](#).
- SimonM finds the [distribution of the maximum price of BTC implied by market prices](#) (cf. [volatility smile](#)), the historical base rates of [SCOTUS accepting a case](#), and the US House of Representatives [disciplining a member](#).
- Matthew_Barnett explains why he wrote a question [calling out bad predictions by journalists and policy analysts](#).
- isinlor [makes the case for predictions about the longer-term future](#): even if Metaculus doesn’t exist, its questions will almost certainly be archived.
- Sylvian comments on [how to ask good questions](#).

In the News

FiveThirtyEight on [why Republicans outperformed polls again](#). Their two hypotheses are that Republicans are losing trust in (strongly left-leaning) institutions, and that college-graduated Republicans might worry about being ostracized for their political views. However, the connection between that and differential nonresponse seems unclear.

Also from FiveThirtyEight: [Ignore What Potential 2024 Presidential Candidates Say. Watch What They Do](#). I found Senator Obama’s [flat out, no-nonsense denial that he would run for president in 2008](#) particularly striking.

[Coles shows off a powerful forecasting engine](#). I was especially surprised by the following paragraph:

Today, 95 percent of the items that you see in the stores are all on automatic ordering. The store does not need to place any orders. Our analytics behind the system looks at the history, looks at the plans, and projects that order demand for the stores that come from all the delivery centres.

I looked into how hard the insurance industry has been hit by COVID. On the one hand, payouts spiked; on the other hand, insurance companies also got more clients. There isn't much hard data, but overall the first effect seems to [dominate around the world](#). [COVID Insurance Coverage One Year Later](#) describes the situation on the US front, explains that policies were ambiguously written and that courts are still deciding whether COVID-19 should be classified as "physical damage" or a "physical alteration".

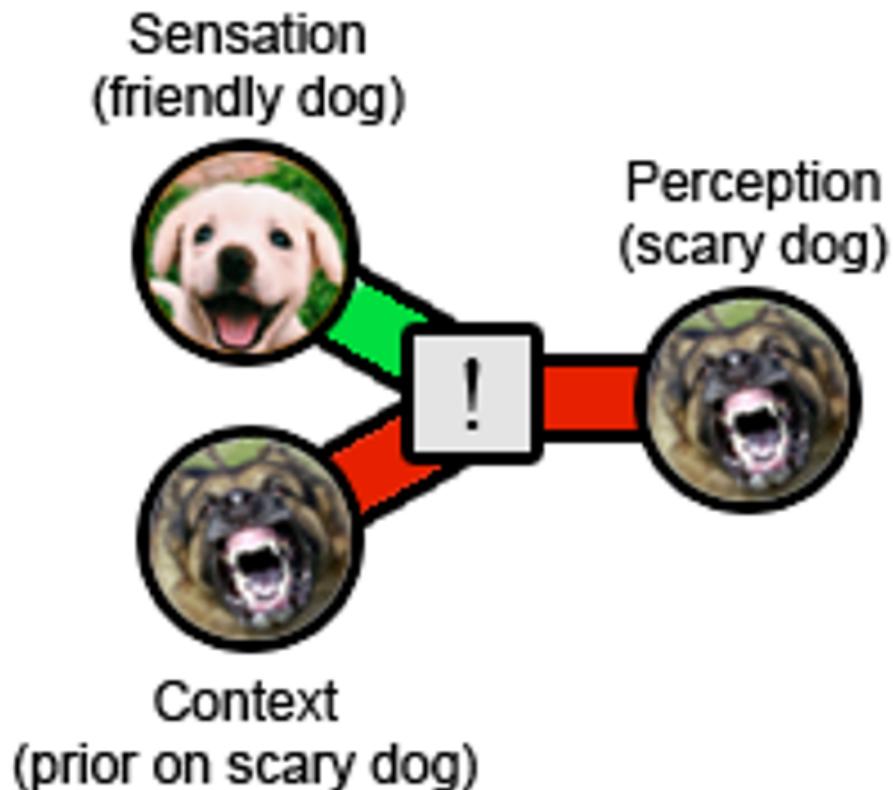
Our next prediction was disheartening, but also pretty obvious: Policyholders would have to sue to secure coverage. The economic impact of this pandemic is larger than anything the world has ever seen, dwarfing other massive loss events like 9/11 or Hurricane Katrina. Paying even a fraction of the claims would bankrupt the insurance industry, so insurers had no choice but to deny every coronavirus-related claim and force policyholders to sue to secure coverage. We got this one right.

[Technology for Forecasting Fish Outbreaks](#) keeps improving. I mention this from time to time, and I may have seen the idea somewhere else, but subsidizing such technology could be a cost-effective intervention to improve fish welfare.

The Association of Bay Area Governments has released a series of demographic, economic, and land-use projections for 2040. The projections are presented [on a sleek webpage](#) as well as in a [more comprehensive pdf](#). The expected error of these projections is difficult to estimate.

Recent Blog Posts

Astral Codex Ten considers [Trapped Priors As A Basic Problem Of Rationality](#). "The raw evidence (the Rottweiler sat calmly wagging its tail) looks promising. But the context is a very strong prior that dogs are terrifying. If the prior is strong enough, it overwhelms the real experience. Result: the Rottweiler was terrifying. Any update you make on the situation will be *in favor of* dogs being terrifying"



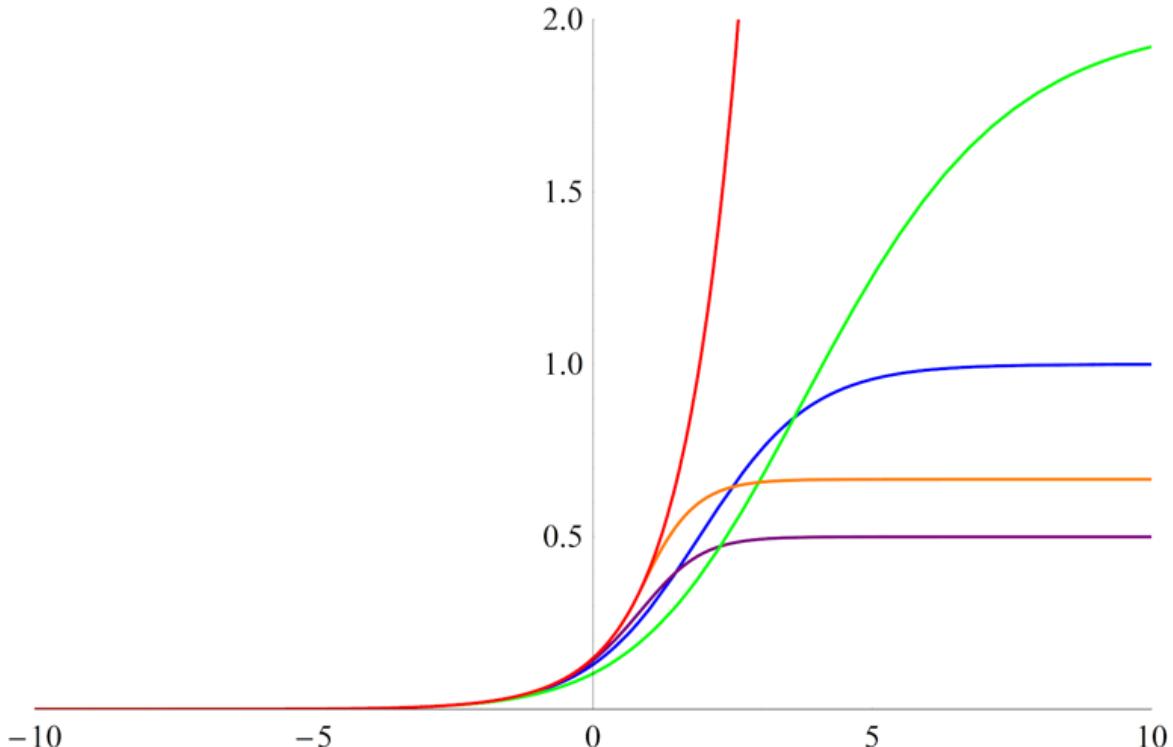
[Astral Codex Ten](#): Priors combine with sensory input to produce a perception of the situation. That perception is then used to produce a new prior.

Discussion on Kelly Betting: [Kelly isn't \(just\) about logarithmic utility](#), [Kelly is \(just\) about logarithmic utility](#), and [A non-logarithmic argument for Kelly](#) (and [this comment](#) which summarizes the last post.)

David Manheim tries to apply accounting principles to forecasting on [Resolutions to the Challenge of Resolving Forecasts](#) and [Systematizing Epistemics: Principles for Resolving Forecasts](#).

deluks917, of previous ["Bet on Biden"](#) fame, has two pieces on the Efficient Market Hypothesis (EMH): [The EMH is False - Specific Strong Evidence](#) and [Violating the EMH - Prediction Markets](#)

[Why sigmoids are so hard to predict](#) makes an argument in terms of the differential equation which produces sigmoids. "The core reason why the turning point and the maximums are so hard to predict from early data [is that] we're not only trying to figure out the parameters of a logistic curve, but the functional form of the dampening function - a dampening function whose effect is insignificant in the early data."



[Why sigmoids are so hard to predict](#): Various growth curves produced by different hard to estimate dampening factors

[Cafebedouin](#), a top Good Judgment Open forecaster who recently ascended into superforecastdom, reviews his predictions for [2020](#).

Niplav looks at [Range and Forecasting Accuracy](#) of questions on PredictionBook and Metaculus. Its results are an instance of [Simpson's paradox](#):

- Questions with a longer range (that is, time between the question being written and the question being resolved) generally receive predictions with a higher accuracy than questions with a shorter range. This might be because they are easier questions, or because they receive higher quality forecasts.
- Predictions made on the same question earlier are generally less accurate than predictions that are made later.

[Star Spangled Gamblers](#) is a political betting blog which mostly covers questions on PredictIt. [Here](#) is a profile piece on whether California's Governor Gavin Newsom will be recalled. The author seems to think that he won't, but that there are many events which would make irrational gamblers push the price higher than it currently is. This thesis is presented together with a solid mechanistic understanding of how California recall elections work and have turned out in the past.

Hard to Categorize

[Forecasting: Principles and Practice](#) is a free online textbook which covers time series forecasting using R.

[Orbit](#) is "a Python package for Bayesian time series modeling and inference" developed by Uber. The [documentation](#) looks reasonably interesting.

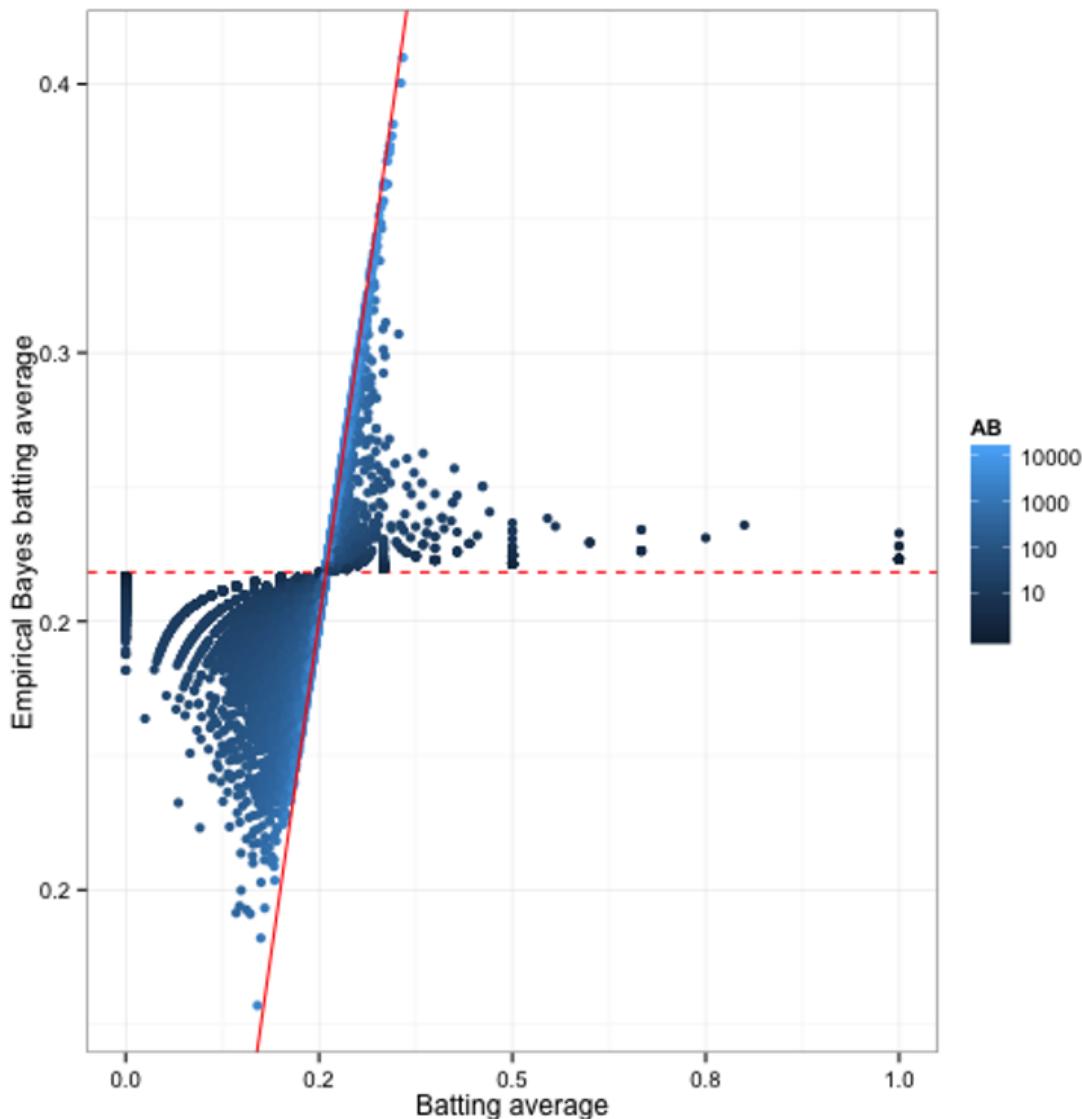
Long Content

OpenPhilanthropy released a report on [outside view perspectives on the likelihood of AGI](#). The report “ignores some of our evidence about when AGI will happen. It restricts itself to *outside view considerations* - those relating to how long analogous developments have taken in the past. It ignores evidence about how good current AI systems are compared to AGI, and how quickly the field of AI is progressing. It does not attempt to give all-things-considered probabilities.”

OpenPhilanthropy asked various academics for feedback. Among other comments, they highlighted the following:

- The importance of [unknown unknowns](#). They would presumably make the prior wider, and could also be incorporated from an outside-view perspective.
- [The assumption of independent and identically distributed trials might be faulty](#). In this case, modelling the path to AGI as a journey of unknown duration ends up giving similar results after an initial period.
- The need for a measure of the robustness of probabilities.
- The observation that AGI is [more of a continuous than a binary problem](#).

[Understanding "empirical Bayes estimation" \(using baseball statistics\)](#): Given two baseball batters, one which has hit 4 out of 10 balls, and another one which has hit 300 out of 1000 balls, which one is, in expectation, better?



[Understanding "empirical Bayes estimation"](#): Batting average before and after adjustment for the number of trials

In a prediction market in which participants Kelly bet, the market price reacts exactly as if updating according to Bayes' Law. See [an introductory blog post](#), and [this paper with a proof](#).

From [The risks of communicating extreme climate forecasts](#):

In a new paper published in the International Journal of Global Warming, Carnegie Mellon University's David Rode and Paul Fischbeck argue that making such forecasts can be counterproductive. "Truly apocalyptic forecasts can only ever be observed in their failure—that is the world did not end as predicted," says Rode, adjunct research faculty with the Carnegie Mellon Electricity Industry Center, "and observing a string of repeated apocalyptic forecast failures can undermine the public's trust in the underlying science."

Fischbeck noted, "from a forecasting perspective, the 'problem' is not only that all of the expired forecasts were wrong, but also that so many of them never admitted to any uncertainty about the date. About 43% of the forecasts in our dataset made no mention of uncertainty."

In some cases, the forecasters were both explicit and certain. For example, Stanford University biologist Paul Ehrlich and British environmental activist Prince Charles are serial failed forecasters, repeatedly expressing high degrees of certainty about apocalyptic climate events.

Rode commented "Ehrlich has made predictions of environmental collapse going back to 1970 that he has described as having 'near certainty'. Prince Charles has similarly warned repeatedly of 'irretrievable ecosystem collapse' if actions were not taken, and when expired, repeated the prediction with a new definitive end date. Their predictions have repeatedly been apocalyptic and highly certain...and so far, they've also been wrong."

[Long-Term Capital Management](#) is a failed hedge fund. In the aftermath of its failure, its manager set up [another hedge fund](#), which also failed in the 2008 crisis, and then a [third one](#), whose current existence is uncertain.

Initially successful with annualized return of over 21% (after fees) in its first year, 43% in the second year and 41% in the third year, in 1998 it lost \$4.6 billion in less than four months due to a combination of high leverage and exposure to the 1997 Asian financial crisis and 1998 Russian financial crisis.

[A Semitechnical Introductory Dialogue on Solomonoff Induction](#) presents, in dialogue form, an idealized way of "how to do good epistemology" if one had infinite computing power.

[Nature study which gives electric shocks to participants when they predict incorrectly](#) finds that "irreducible subjective uncertainty" is very predictive of stress.

On each trial, a stimulus (rock A or rock B) was presented and participants were asked to predict whether or not there was a snake underneath (snake or no snake). Each time a snake was presented, participants received a painful electric shock to the hand.

Pupil diameter and skin conductance provided established measures of activity in the autonomic nervous system, a key effector of acute stress responses.

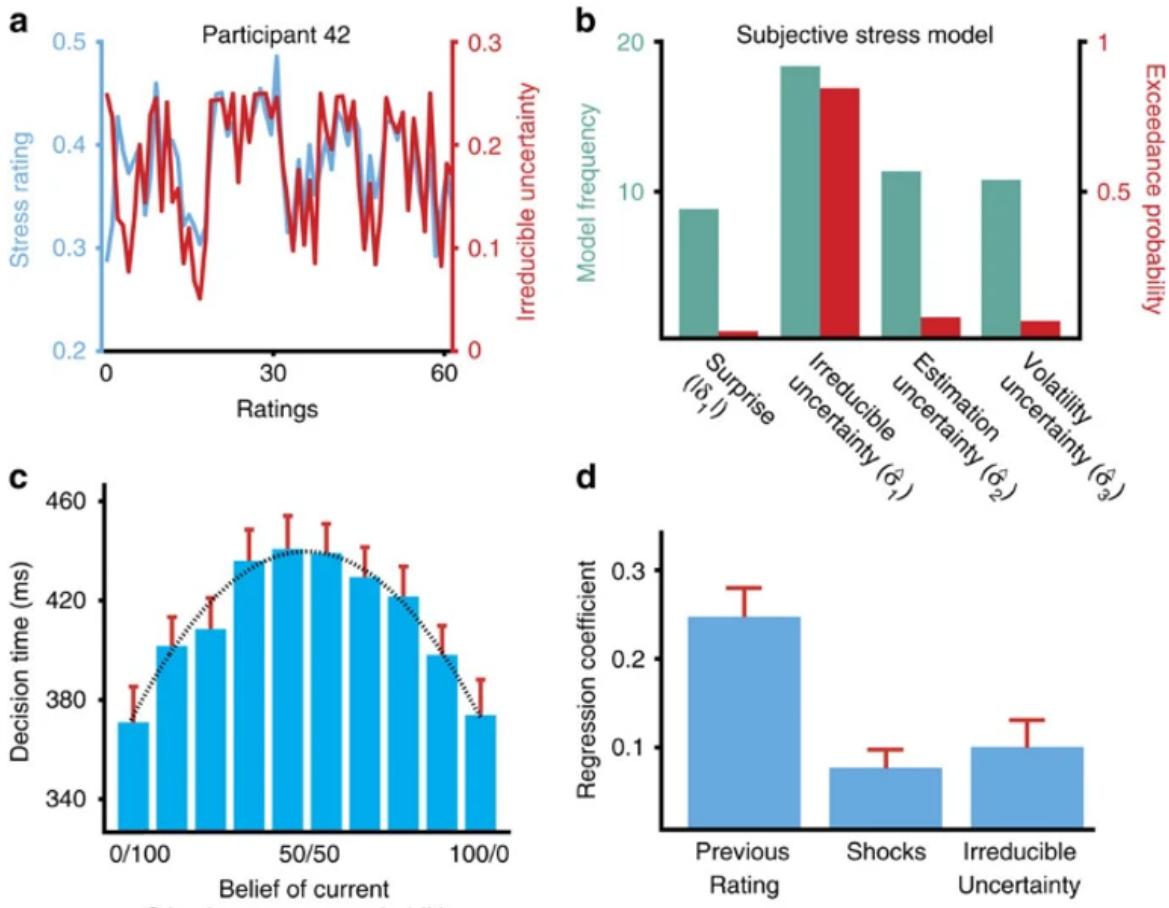
We found that all three were predicted by subjective irreducible uncertainty. We further examined interindividual variance in the degree of coupling between uncertainty and stress responses, which we related to the ability of participants to learn in an uncertain dynamic environment. Unpredictable aversive threat induces stress.

The probabilistic mapping from stimulus (rock) to outcome (snake) shifted over the course of the experiment (Fig. 1c), requiring participants to track this relationship over time. When an outcome was revealed, the presence of a snake was deterministically associated with an electric shock delivered to the back of the left hand. Over the course of 320 trials, the probabilistic mapping between stimuli and outcomes changed every 26–38 trials, requiring participants to maintain and update their beliefs about the probability of a snake being under either rock.

Shocks and irreducible uncertainty both predicted subjective stress ratings (single-sample t-tests, $P<0.001$; $P=0.0024$).

As predicted, participants reported being most stressed when they believed the current state was high in irreducible uncertainty.

Subjective irreducible uncertainty is highest in our task when the subject's estimated probability of a shock is 50%, corresponding to a situation where the environment is utterly unpredictable, and maximal in entropy.



[Computations of uncertainty mediate acute stress responses in humans: Fig. 4.](#)

Note to the future: All links are added automatically to the Internet Archive. In case of link rot, go [there](#) and input the dead link.

Value is not created in the production of the forecast, but in the deployment of plans and actions that follow.

Source: [Kitsch article about how to improve forecasting.](#)

Forecasting Newsletter: April 2021

Highlights

- Polymarket is being attacked by “sandwiching” bots
- Metaculus launches “[Forecasting Causes](#)”
- In [Reflective Bayesianism](#), Abram Demski outlines questionable and implicit assumptions which Bayesians make.

Index

- Prediction Markets & Forecasting Platforms
- In The News
- Blog Posts
- Long Content
- Hard To Categorize

Sign up [here](#) or browse past newsletters [here](#).

Prediction Markets & Forecasting Platforms

Polymarket

After the demonstrable success of Polymarket (and, to a lesser extent, Augur) in attracting volume to their platforms, many imitators have popped up on the crypto scene. None of them are functional yet, but I thought I'd mention them, in order from least to most scammy:

- [PolkaMarkets](#) is aiming for an August release date, and has recently begun [testing its MVP](#).
- [Hedgehog Markets](#)'s schtick is to be implemented in an up-and-coming blockchain, Solana. The functioning markets on their webpage are currently using test money.
- [Totem](#) has an interesting scheme where predictions are non-punitive. That is, people who predict incorrectly won't lose money, they will merely win less. However, after reading [their whitepaper](#), it is not exactly clear where the money for rewards will come from. It is being built on top of the Binance chain, [a centralised exchange with a centralized coin](#), which I find unappealing.
- [Polars](#) and [PredictX](#) are also built on top of the Binance chain. I would characterize them as money grabs. That is, they are attempting to raise money to do something like Polymarket without a clear plan for how they would be superior.

While Totem feels scammy, there are interesting possibilities related to its core idea. For example, external actors could subsidize the market. Another alternative could be to stake the bet amounts on a financial instrument which provides returns, like Uniswap, while waiting for resolution. In other words, one could bet the interest, not the principal.

While imitator projects go through the design and test stages, Polymarket has been dealing with a new real world problem: “sandwiching”. Here is how it works on a high level: a bot detects transactions before miners process them, and profits from that information at the expense of the user.

To understand it on a more detailed level, it is first necessary to explain several details about Polymarket's architecture. Polymarket uses an Automated Market Maker design (as opposed

to an order book). This means that users trade with liquidity providers, who take bets on both sides, rather than with other users. This enables users to make bets at any time, even when there isn't somebody willing to take the other side. Liquidity providers take a small fee, and are exposed to risk if too many people want to bet on the correct side at once. To reduce that risk, liquidity providers change the odds with each bet—the more people bet in the same direction, the more the odds change. This happens on a blockchain, so miners—the players who add transactions to the record of transactions—need to be able to see transactions, such as bets. But, this means that other players, and in particular bots, are able to see them too. The miners prioritize transactions based on how much users are willing to pay, so a bot which wants to jump the line can pay a little bit more to do so.

Now, when a user is going to make a bet, a bot can buy $\$b$ worth of a contract at price X (equivalent to betting at odds $X : (1 - X)$), which moves its price to $X + \text{slippage}(b)$. The user then buys $\$u$ worth of a contract at $X + \text{slippage}(b)$, and moves the price to $X + \text{slippage}(b) + \text{slippage}(u)$. The bot then sells its shares at $X + \text{slippage}(b) + \text{slippage}(u)$, moving the price to approximately $X + \text{slippage}(u)$. The bot bought at $X + \text{slippage}(b)$, and sold at $X + \text{slippage}(b) + \text{slippage}(u)$, so it made a profit of approximately $\text{slippage}(u)$ per share, at the expense of making the user buy at a more expensive price (at $X + \text{slippage}(b)$, instead of at X). Crucially, because the bot also pays fees to liquidity providers for its two transactions, the attack is only profitable if $\text{slippage}(u)$ is large enough. For example, Polymarket has 2% fees, so in a 50% contract, the attack is only profitable if the user moves the price by more than 4%, i.e., if $\text{slippage}(u)$ is bigger than 4%. This calculation changes somewhat when the price moves away from 50%.

Polymarket has implemented "slippage protection", which solves a part of this problem. In particular, it detects that a bot (or another user) has moved the price from the expected X to $X + \text{slippage}(\text{someone})$, and halts the trade if this is the case. But, for this protection to be effective, the user then has to refuse to buy at the new price of $X + \text{slippage}(\text{someone})$ after the trade has failed. Polymarket could improve on this by dividing a trade into small, unpredictably-sized chunks randomly delayed in time. Then it wouldn't be profitable to sandwich each trade individually, and it would also be difficult to sandwich the whole sequence—a bot wouldn't be able to know whether a trade begins a sequence. Still, there is currently a particularly annoying bot which appears to be sandwiching all trades, even when it isn't be profitable.

Background reading: [Ethereum is a Dark Forest; I got “sandwich attacked” on my transaction](#)

Avraham Eisenberg writes [Tales from Prediction Markets](#), taking place on Polymarket. It features market manipulation, bold exploits, and cautionary tales. Polymarket itself also has announced a ["Microgrants"](#) program for people to work on projects related to the platform.

Metaculus

Metaculus has announced ["Forecasting Causes"](#), an initiative to connect non-profit organizations with forecasters. They are starting with a [tournament on alternative meat](#) (see

also: [more commentary](#)), a part of their [Feeding Humanity](#) cause area, and a [tournament on COVID-19 in Virginia](#), as part of their [Healthy Communities](#) cause.

Background reading: Long Term Future Fund grants to Metaculus, back in [2019](#) and [2020](#).

Metaculus is [revamping](#) a part of its incentive designs mechanism. They are also [hiring](#) for Junior Designer, Full-Stack Developer, and Public Policy Data Scientist positions.

[SimonM](#) kindly curates the top comments from Metaculus this past April. They are:

- [GlobalGuessing](#) and [PeterHurford](#) lay out the chances of an Olympic boycott in 2022.
- [EvanHarper](#) points out the base rate for KIA is much lower in modern wars.
- Metaculus users have fun with a [pseudo-Keynesian Beauty Contest](#).
- SimonM and Charles discuss [Jannik Sinner's prospects](#).
- [elifand_ought and liconstan](#) have very different forecasts for whether or not the meat industry will put out an anti-plant-based-food ad.
- [onlyasith's](#) forensic analysis of a Brandon Adams podcast is part of his model for when Nate Silver's next book will be out.
- [fianxu](#) points out jury instructions don't apply to forecasters.
- [haukurth](#) notes the community forecast was well calibrated during the Chauvin trial.
- [SimonM](#) points out people don't update their forecasts as often as they should.
- [clearthis](#)'s comment causes people to notice an old question about Trump family indictments, and the Metaculus community prediction moves up by 30%.

SimonM has also added more functionality to the Metaculus Extras site, including a "[movers and shakers](#)" page. It lists the Metaculus questions that moved the most in the last week.

In the News

The Anticipation Hub is a loose association of organizations trying to anticipate catastrophes and mitigate them before they happen. Based on the contents of their newsletter, they seem to be [fairly active](#), particularly around the topics of floods, tropical storms, and inter-institutional cooperation. For instance, they have a [postmortem](#) of the actions they took before a severe tropical storm "Chalane" hit Mozambique. They find that the value of their preparation decisions was not as high as they had hoped for.

[Charismatics issue 'prophetic standards' to address false Trump prophecies \(original source\)](#). After an embarrassing number of wrong prophecies and bungled predictions about the 2020 election, a group of [charismatic Christian](#) leaders have released a four-page statement of "prophetic standards" to help correct abuses in the movement". In particular, they call on those who have made false prophecies to apologize.

[How spooks are turning to superforecasting in the Cosmic Bazaar](#): The Economist mentions a few details about the "Cosmic Bazaar", a forecasting tournament organised by the British government.

Since the launch of the website in April 2020, more than 10,000 forecasts have been made by 1,300 forecasters, drawn from across 41 government departments and several allied countries. The site has around 200 regular forecasters monthly, who must draw only on publicly available information to tackle 30-40 questions live at any time. Cosmic Bazaar represents the gamification of intelligence. Users are ranked by a brutally simple measure: the accuracy of their predictions.

Facebook releases a research paper: "Large-scale forecasting: Self-supervised learning framework for hyperparameter tuning" ([blog post](#), [arxiv preprint](#))

Forecasting is one of the core data science and machine learning tasks we perform at Facebook, so providing fast, reliable, and accurate forecasting results with large amounts

of time series data is important for our business.

We empirically evaluated our algorithms on both internal and external data sets, and obtained similar conclusions. SSL frameworks can dramatically improve the efficiency of model selection and hyperparameter tuning, reducing running time by 6-20x with comparable forecasting accuracy.

This approach is independent of specific forecasting models and algorithms

Goldman Sachs [downgrades](#) India's growth forecast as Covid cases spike.

The European Centre for Disease Prevention and Control (ECDC) has run a new forecasting hub for about a month now ([announcement](#), [link to the hub](#)). It's unclear whether any decision-makers are influenced by its predictions.

[Draft EU legislation on AI is proposing to regulate the use of Bayesian estimation](#) (see also some commentary [here](#) [here](#), and [here](#)). Some commenters opine that the proposal is business as usual,—draft legalese that will get refined and clarified in a future review. Other commenters view it as another example of the EU's incompetence.

Blog Posts

[The Kelly Criterion Kinda Sucks](#). An experienced PredictIt bettor points out that the Kelly criterion doesn't offer guidance in cases when one can bet on more than one event at more than one point in time.

Scott Alexander, of AstralCodexTen, publishes a [list of 75 public predictions for 2021](#). Zvi and [SimonM](#) discuss some aspects of it on LessWrong. I have also added them to [Foretold](#), in case people want to forecast on them, and to [Metaforecast](#), which makes them searchable using a nicer interface.

[The die is forecast](#) is a "Computational Social Science and Political Event Analysis" blog by the people behind [CoupCast](#), and the [Rulers, Elections and Irregular Governance \(REIGN\)](#) dataset. I appreciate their rigorous, base-rate-driven analysis of [election violence risk](#) in upcoming elections around the world.

[Global Guessing](#) is a geopolitics forecasting blog. The authors take questions from various forecasting platforms, chiefly Metaculus, and analyze them in depth in public. For instance, see [this post](#) on the chances of a boycott of the 2022 Winter Olympics in China, the chances of a North Korea ICBM test, and the possible origins of COVID-19.

[Stephen S. Roach](#) is a senior fellow at Yale. In [My Worst Forecasting Mistake](#), he looks back at one of his failed predictions:

Attempting to predict interest rates was my least favorite part of the job. With good reason. I remember walking into the old Morgan Stanley investment banking meeting room and seeing a chart of my predecessor's bond market forecast sitting upside down on the floor. I was determined to avoid that fate. When my favorite bond trader started calling me "dart man," I made an executive decision to disengage and hire an interest-rate strategist. Survival of the fittest, I guess.

I went on to stress that the nascent recovery was likely to be aborted by a relapse, as had occurred in eight of the preceding 11 recessions since the end of World War II. A few months later, taking comfort from some economic indicators that had broken my way, I committed the most egregious forecasting sin of all: giving a date. I actually wrote that the coming double-dip was likely to occur by mid-2021.

In the end, the confluence of science, politics, and the indomitable human spirit left my out-of-consensus double-dip call in tatters. It wasn't my first forecasting mistake, but it is probably the most glaring. Mea culpa is an understatement. Back to the ivory tower.

He has it right that for him as an individual, it was the wrong move to make a quantifiable prediction. But, the system as a whole wants for people like him to make quantifiable predictions and be weeded out by more accurate forecasters.

In [The Johnson & Johnson Pause Shows The System Is Working](#), FiveThirtyEight claims that stopping vaccination because of one death and six cases of a rare type of blood clots among 6.8 million doses administered was somehow a positive development.



[Some "This is fine" dog meme variations](#), @LinchZhang, 2021

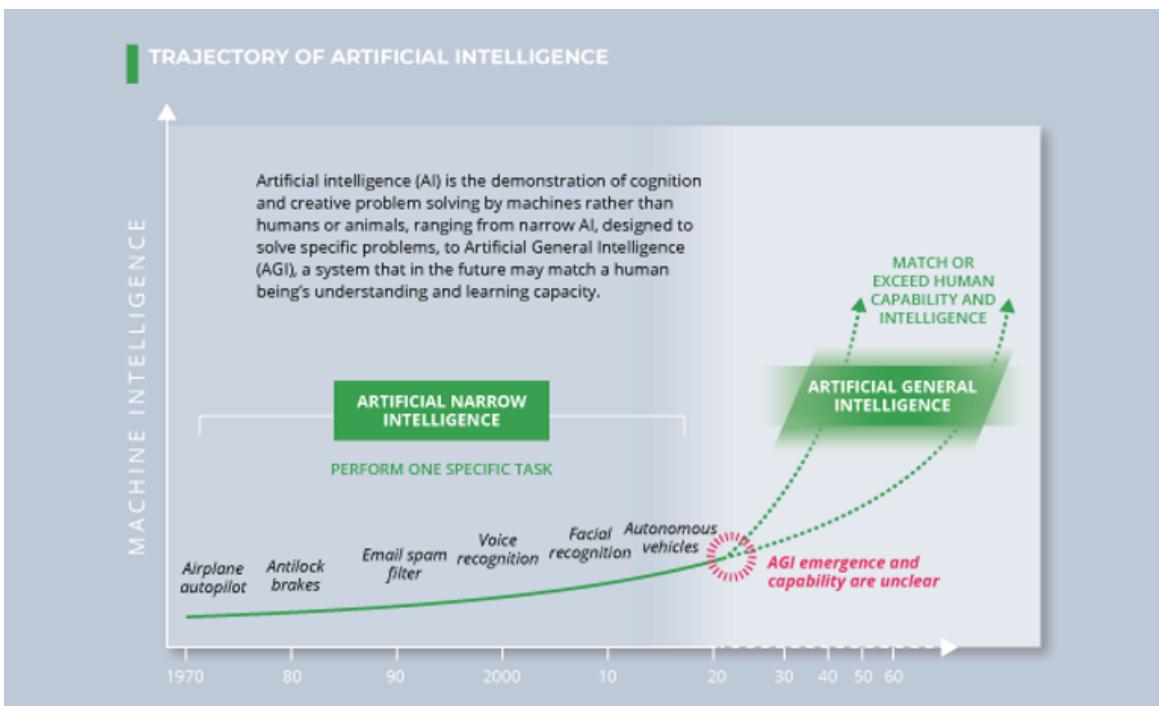
Abram Demski writes [Reflective Bayesianism](#), which outlines implicit and questionable assumptions that strict Bayesians have to make (cf. [Radical Probabilism](#)).

[Probability theory and logical induction as lenses](#) talks about how both probability theory and logical induction are both "lenses for looking at the real-world phenomena of machines that quantify their uncertainty in their beliefs".

Long Content

The Office of the Director of National Intelligence (ODNI) of the USA has released a "[Global Trends 2040](#)" report. In ancient history, the Bush administration established this office in the aftermath of the US intelligence community's failure to aggregate information available in disjoint agencies to predict the 2001 attacks (cf. [Wikipedia](#)). Then in 2006, [IARPA](#) was formed as an organization within the ODNI, and went on to organize the [ACE program](#), which explored which kind of setups produced the most accurate probabilities, and found out that pre-selected superforecasters from the general population working in groups did best, or at least better than intelligence analysts with access to classified information.

Because of this history and background of its office, I was expecting the report to have some probabilistic estimates, but the report is instead structured around a set of scenarios, rather than around quantified predictions. However, the report does include interesting observations around structural technology forces:



[Global Trends 2040, Structural Forces: Technology](#), Office of the Director of National Intelligence, 2021.

Timelines Shrinking. The time to develop, deploy, mature, and then retire technologies is moving from decades to years and sometimes faster. Multiple actors, including corporations and states, at the forefront of emerging technology may deploy and exploit a new technology before others get off the starting blocks. Those trying to catch up, especially in developing countries, may be increasingly forced to choose technologies before the implications of those choices are fully understood, risking investment in technological dead ends or falling hopelessly behind. Planned economies may be able to react faster to emerging technology developments, potentially at the cost of reduced technological diversity and efficiency.

AI is the demonstration of cognition and creative problem solving by machines rather than humans or animals, ranging from narrow AI, designed to solve specific problems, to Artificial General Intelligence, a system that in the future may match or exceed a human being's understanding and learning capacity. By 2040, AI applications, in combination with other technologies, will benefit almost every aspect of life, including improved healthcare, safer and more efficient transportation, personalized education, improved software for everyday tasks, and increased agricultural crop yields. Political and business leaders worldwide are seeking global talent and are pouring resources into developing AI, hoping to be among the first to use it to reshape societies, economies, and even war. Enabled by concurrent increases in high-quality data, computing capability, and high-speed communication links, AI will challenge leaders to keep pace and reap the benefits while mitigating harmful effects, such as threats to privacy and liberty.

Existential Risks. Technological advances may increase the number of existential threats; threats that could damage life on a global scale challenge our ability to imagine and comprehend their potential scope and scale, and they require the development of resilient strategies to survive. Technology plays a role in both generating these existential risks and in mitigating them. Anthropomorphic risks include runaway AI, engineered pandemics, nanotechnology weapons, or nuclear war. Such low-probability, high-impact events are difficult to forecast and expensive to prepare for, but identifying

potential risks and developing mitigation strategies in advance can provide some resilience to exogenous shocks.

In contrast to the approach taken by the report above, [Keeping Score: A New Approach to Geopolitical Forecasting](#), by Perry World House, calls for more quantified predictions. See also [this twitter thread](#) by @MWStory.

The [Survey of Professional Forecasters](#) is "the oldest quarterly survey of macroeconomic forecasts in the United States. The survey began in 1968 and was conducted by the American Statistical Association and the National Bureau of Economic Research. The Federal Reserve Bank of Philadelphia took over the survey in 1990." Its last forecasts, for Q1 2020, can be found [here](#).

Robin Hanson does some Robin-Hansoning, and proposes "[Shoulda-Listened Futures](#)", a system in which people who think they are unfairly ignored by mainstream science could pay for a future evaluation of their work and for an immediate prediction about that evaluation.

These "world shoulda listened to me" customers might pay to have some of their works evaluated by posterity. For example, for every \$1 saved now that gains a 3% real rate of return, \$19 in real assets are available in a century to pay historians for evaluations. At a 6% rate of return (or 3% for 2 centuries), that's \$339. Furthermore, if future historians needed only to randomly evaluate 1% of the works assigned them, then if malcontents paid \$10 per work to be maybe evaluated, historians could spend \$20K (or \$339K) per work they evaluate. Considering all the added knowledge and tools to which future historians may have access, that seems enough to do a substantial evaluation, especially if they evaluate several related works at the same time.

Given a substantial chance (1% will do) that a work might be evaluated by historians in a century or two, we could then create (conditional) prediction markets now estimating those future evaluations. So a customer might pay their \$20 now, and get an immediate prediction market estimate of that future evaluation for their work. That \$20 might pay \$10 for the (chance of a) future evaluation and another \$10 to establish and subsidize a prediction market over the coming centuries until resolution.

Note to the future: All links are added automatically to the Internet Archive. In case of link rot, go [there](#) and input the dead link.

I'm not disagreeing with your emotions or your politics. I'm disagreeing with your numbers.

[@LinchZhang](#), 2021.

Forecasting Newsletter: May 2021

Highlights

- Misha Yagudin creates a [webpage](#) to get one's calibration chart for Good Judgment Open and CSET-Foretell
- Hypermind experiments with new methods of eliciting, incentivizing and scoring long-range forecasts.
- Augur launches [Augur Turbo](#) on Polygon, but gets very little trade volume

Index

- Prediction Markets & Forecasting Platforms
- In The News
- Blog Posts
- Papers
- Hard to Categorize
- Long Content

Sign up [here](#) or browse past newsletters [here](#).

Prediction Markets & Forecasting Platforms

Augur

Augur has launched [Augur Turbo](#) on Polygon, the same second-layer chain which also hosts Polymarket. The website to access it is hosted on IPFS, hence the long address when one accesses it through a [portal](#).

To unpack the tech stack:

- [Augur](#) is a decentralized prediction market protocol. Its strength lies on its robust decentralized question resolution mechanism: resolutions can be disputed until a consensus is reached.
- [Polygon](#) (previously Matic) is a parallel blockchain for Ethereum. It offers contract compatibility, so that contracts written for the main Ethereum blockchain can also run on Polygon. Moving assets from the main Ethereum blockchain to Polygon is initially somewhat cumbersome. But once inside Polygon, transaction fees are much cheaper.
- [Chainlink](#) is an oracle service. Most of the time, it will work and it will work fast. The rest of the time, resolution can be disputed and Augur's protocol can be used instead. Chainlink claims to be decentralized, but appears [not](#) to fully [be so](#).
- [IPFS](#) is a censor-proof, decentralized alternative to the [http protocol](#). To a first approximation, files in the network are [hashed](#) using the [SHA-256](#) algorithm, and accessed using their hash. However, this turns out to be [more complicated](#) in practice.

So far, Augur Turbo merely has NBA markets, with low volume, and low liquidity.

Augur Turbo [webpage](#). Screenshot taken the 1st of June, 2021. Notice that the market with the highest total volume has a mere \$100 in trade volume.

From Augur's early history, [Is Augur Being Gamed](#) explains what exactly [Poyo](#) did to profit from the creation of invalid markets. Back in the day, Augur didn't have a tradable "invalid" resolution. Instead, invalid or ambiguous markets were resolved 50-50. This allowed Poyo to profit by creating markets he knew would resolve as invalid, and then buying the cheaper side.

CSET-Foretell

[CSET-Foretell](#) has added the ability for users to suggest questions.

Good Judgment training for "Foretell Pros"—the best scoring forecasters during CSET-Foretell's first season—continues. Because Foretell Pros might get culled if they perform worse than the crowd, and because their score is proportional to their difference from the crowd, they have an incentive not to share information. When this was pointed out, CSET-Foretell answered with an impassioned appeal to the better angels of our nature. It seems it worked to some extent, and participants are sharing more of their reasoning within the platform and community.

Good Judgment Inc/Good Judgment Open

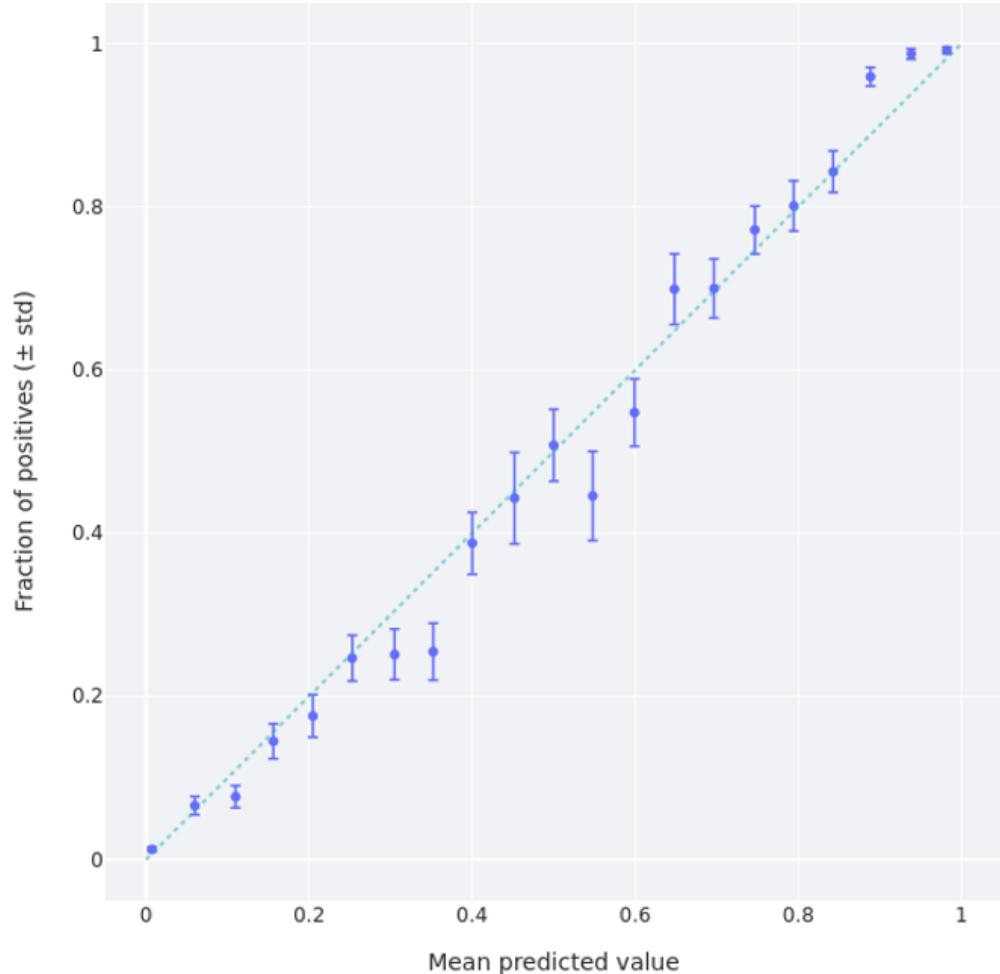
Per the [Good Judgment Open Newsletter](#), product lead and Superforecaster Luis Enrique Urtubey De Césaris has some openings in his office hours coming up on June 11th and 18th. The contact email provided is beta@goodjudgment.com, no schedule is given.

The Financial Times [reports](#) on Superforecaster predictions for "When will the number of doses administered globally reach 5 billion?"

How to find out your calibration on Good Judgment Open and CSET-Foretell

Misha Yagudin, friend of the newsletter and fellow Samotsvety Forecasting team member on CSET-Foretell, has programmed a [site](#) which allows users to get their calibration chart for their predictions on Good Judgment Open and other Cultivate Labs platforms, such as CSET-

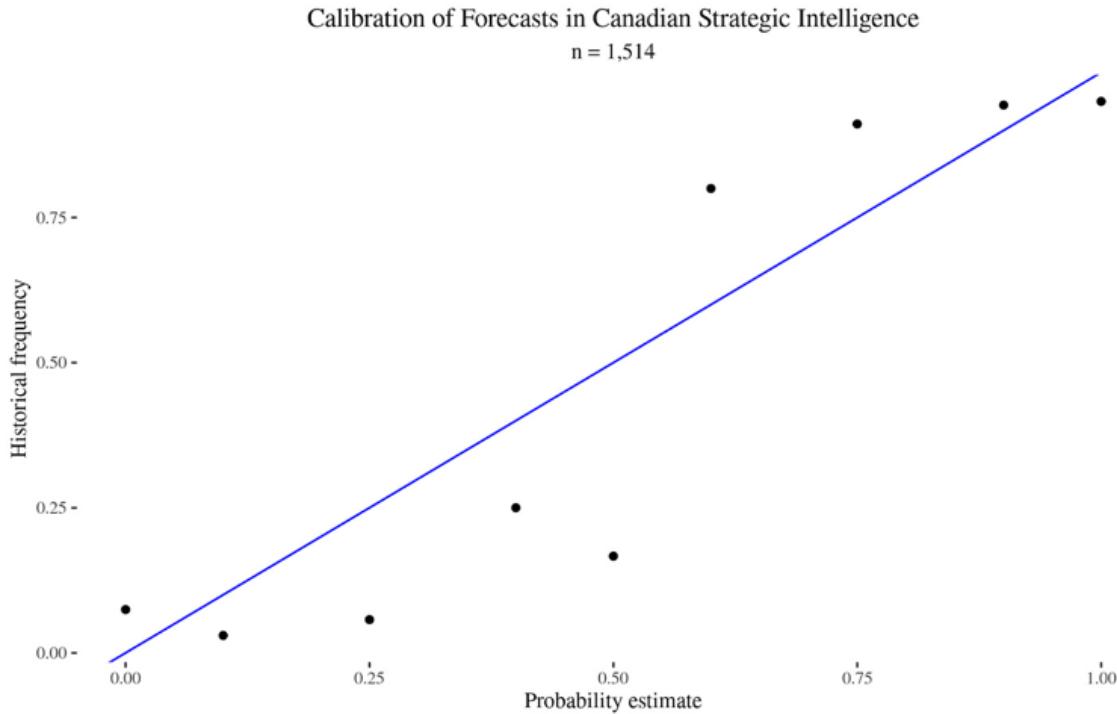
Foretell. See [this one-minute video](#) for how to use it. Using [Misha's site](#), my calibration chart looks as follows:



My calibration chart for Good Judgment Open as of the 1st of July 2021, based on 2337 forecasts on 106 resolved questions.

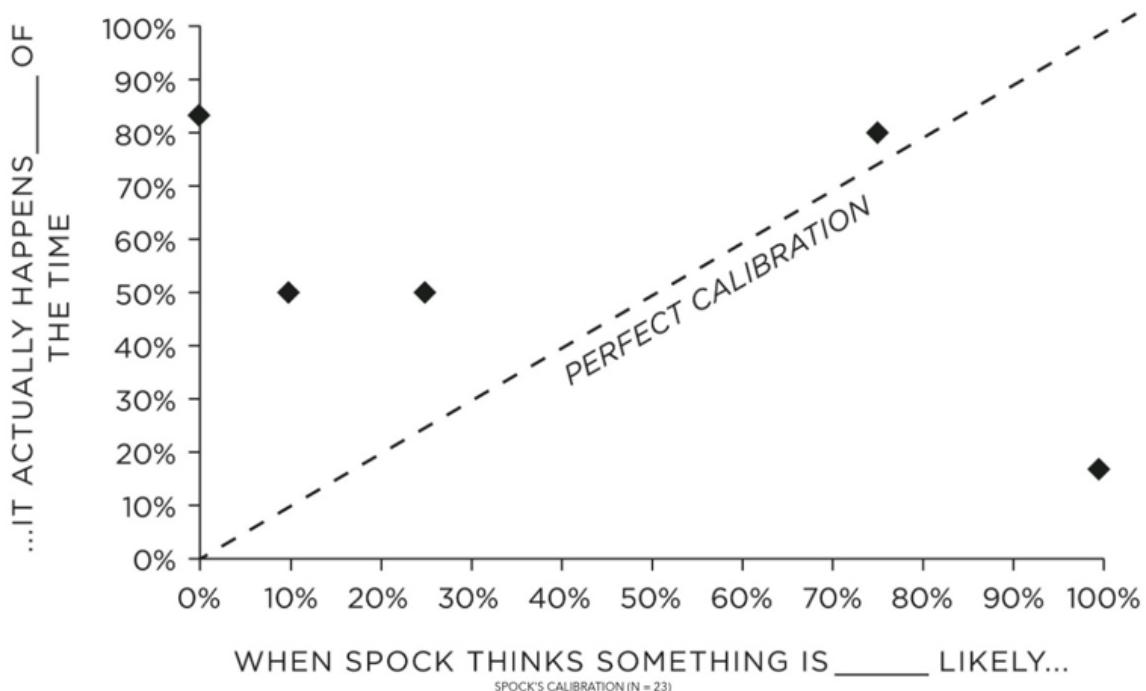
This means that I'm under-confident around the 15% (resp. 85%) level. I know why this is: I was assigning a 15% chance to questions which gave me the feeling that "this is most likely not the case, but I'm not completely sure." As it turns out, the kinds of questions on Good Judgment Open which generate that feeling instead happen around 10% of the time.

For comparison, here is the historical calibration of Canadian strategic intelligence forecasts, which I calculated using [this dataset](#):



[Accuracy of forecasts in \(Canadian\) strategic intelligence](#), processed [in R](#)

And here is the calibration of [Spock](#), from Julia Galef's *The Scout Mindset* (h/t Gavin Leech, Michał Dubrawski):



Spock's calibration chart, taken from Julia Galef's *The Scout Mindset*, p.77.

Here are some hypothesis about why Spock's calibration is skewed:

- The [anthropic principle](#). Even if Spock was perfectly calibrated in his estimates that a given suicidal maneuver had a 10% chance of survival, we would only hear about the

adventures of the Enterprise crew if they survive after the fact. Or, in other words, consider that the adventures of the Enterprise and the other ships are selected to be the most interesting ones out of a whole interstellar civilization. So we can't infer from the fact that Kirk's crew won a lottery ticket that the probability of winning was a priori high. Consider also that (potentially infinite) [mirror universes](#) are canon within the Star Trek universe. This means that selection effects may be arbitrarily pronounced, and we might just be observing the thinner and thinner slices of probability mass where Kirk tries something improbable and survives. See also [The Hero With A Thousand Chances](#).

- [Plot armor](#). By virtue of being a straight-faced foil to Kirk, Spock cannot have "we are in a story" in his hypothesis space. But maybe Kirk does, and this would explain why Kirk's implied probability assessments are better than Spock's. See also [Reflective Bayesianism: The world is in our hypothesis-space](#). A variation of this hypothesis would be the existence of [a God](#) within the Star Trek universe.
- Spock's deep seated anti-human speciesism. If Spock truly was as rational as advertised, he should have noticed that his calibration was way off. Then, if nothing else, he could apply a calibration adjustment. I find this a parsimonious explanation, given that Spock is often shown as deeply conflicted about his mixed human-Vulcan heritage. In particular, having been bullied on Vulcan as a kid because of that mixed heritage might have led to a stunted psyche in this regard.

Hypermind

Hypermind has been experimenting with new methods of eliciting, incentivizing and scoring long-range forecasts. The first mechanism consists of "drip rewards." In short, if you want to get predictions about an event in 2030, you could try to promise forecasters a reward in 2030. But they might not find it very motivating. Instead, you could ask the same question each year (2021, 2022,...) until 2030, and reward forecasters according to how much their prediction one year resembles the crowd's predictions in the next year.

[Here](#) is a summary of the mechanism which includes more twists, such as making the reward time random, and increasing rewards as resolution time approaches. Hypermind is trying out this method for [predicting COVID-19 vaccinations by 2029](#), with a price pool of \$30,000.

Comments from colleagues centered around the fact that Hypermind wants to patent the method, but there is plenty of prior art. For instance, in machine learning a similar idea is known as [Temporal difference learning](#) (h/t Misha Yagudin.)

The second method is more speculative. Forecasters make an object-level prediction, and a meta-prediction on what the crowd prediction will be. Then, forecaster predictions are adjusted—based on the meta-predictions—to increase the probability of "surprisingly popular" predictions. See Wikipedia on the [Surprisingly popular](#) method for a simplified example.

There are four groups of people:

A: "Philadelphia is the capital of Pennsylvania, and others will agree." (This group answers yes/yes.)

B: "Philadelphia is the capital of Pennsylvania, but most others won't know that". (This group answers yes/no.)

C: "Philadelphia is not the capital of Pennsylvania, and others will agree." (This group answers no/no.)

D: "Philadelphia is not the capital of Pennsylvania, but most others won't know that." (This group answers no/yes.)

This technique causes groups A and C to be eliminated from consideration and measures the difference in size between groups B and D.

Both groups B and D think they know something other people don't, but B is wrong and D is right. In cases where people feel like they have "inside" knowledge, it's more often the case that it's because they are correct and knowledgeable (group D), not because they are misled (group B).

The [paper from which this method comes](#) considered forecasts on discrete bins. Per its [public writeup](#), Hypermind applies this method to predicting continuous distributions by dividing continuous distributions into discrete bins *and then ignoring bins with probabilities below 5%!*:

A second approximation is the minimum probability required to consider that a forecaster has a "non-zero" probability in a particular bin. This goes to the core of equation (2) which computes the prediction-normalized forecast based on which participants have a forecast or a meta forecast in each bin. After some experimentation, we settled on a rather aggressive threshold of .05 probability below which a forecaster is not considered to have forecasted or meta-forecasted that particular bin.

Hypermind tried this method for predicting [the state of AI in 2030](#) where ignoring events which have a lower than 5% probability seems like a particularly bad idea, given that those events might be particularly impactful.

Metaculus

GlobalGuessing [interviews Gaia Dempsey](#), Metaculus' CEO, and continues [analyzing Metaculus questions](#).

[SimonM](#) kindly curated the top comments from Metaculus this past May. They are:

- [cd](#) argues for uncertainty in the New York mayoral race.
- [EvanHarper](#) reports that Andrew Yang is no longer leading on the New York mayoral race polls.
- SimonM isn't very optimistic about [100m times at the Olympics](#), and notices that [longevity research has cashflows tied to crypto](#).
- [niplav](#) uses the "[Musk forecast correction factor](#)" to predict Starlink internet availability.
- A [lab-leak hypothesis question](#) resolves ambiguously, which leads to discussion ([1](#), [2](#), [3](#).)
- [fianxu](#) (Gaia Dempsey) gives the lowdown on whether a Metaculus mobile app is coming (it isn't.)

A tournament on Virginia COVID-19 cases was also covered by a [quaint local Virginian newspaper](#).

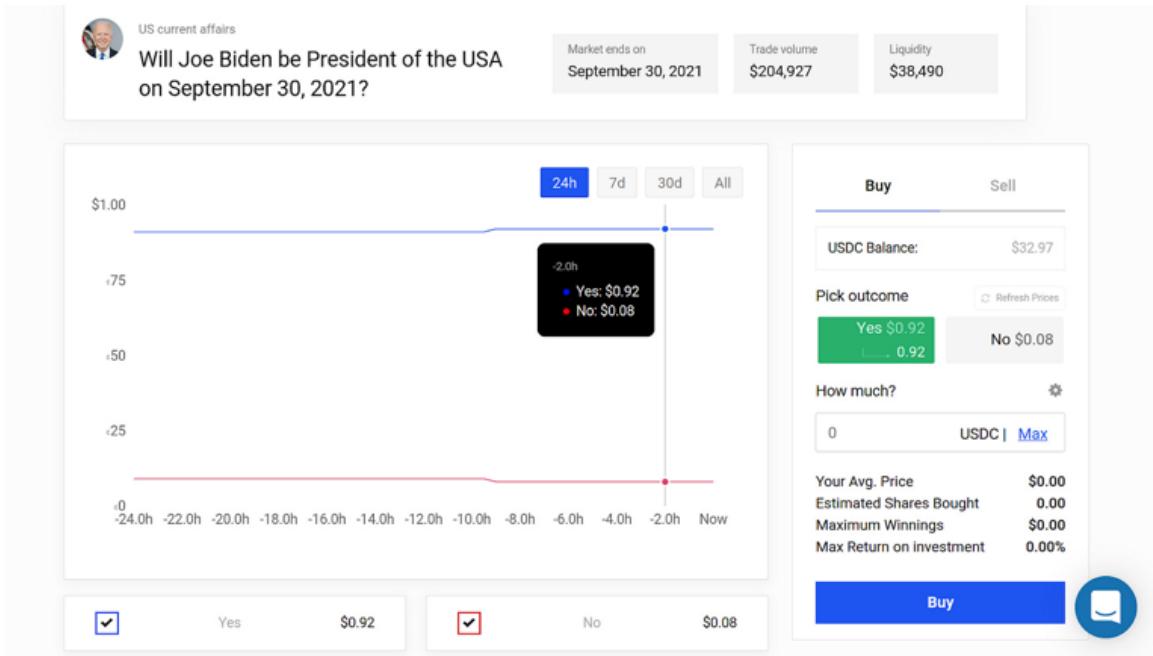
Polymarket

Polymarket featured plenty of markets about [NBA playoffs](#). They also [sponsored](#) GM Hiraru Nakamura's Twitch stream throughout the #FTXCryptoCup, a chess tournament organized by [FTX](#). As part of their sponsorship, they gave away \$20 to 500 new users; it seems like the [link](#) is still up.

Polymarket's microgrants program spawned [Polystats](#), which displays statistics about markets. The site might make it easier for liquidity providers to choose where to stake their funds, and competes with an earlier site, [PolymarketWhales](#).

"Sandwiching" bots, covered in the [previous edition of this newsletter](#), continue to be an annoyance.

As for markets, [Will the 2021 Tokyo Olympics take place?](#) is sitting at ~81% (~77% on [FTX](#)) (!), and [Will Joe Biden be President of the USA on September 30, 2021?](#) is currently sitting at ~92% (!?).



Polymarket: [Will Joe Biden be President of the USA on September 30, 2021?](#). Screenshot taken on the 1st of July 2021.

PredictIt

[Old Bull TV](#) is a Youtube Channel which covers PredictIt markets. Their episode [When PredictIt Met Kevin](#) presents the case of [Kevin Paffrath](#), a random influencer with [1.63M](#) [Youtube followers](#) who [got his followers and associates](#) to buy his shares for the [Who will be the governor of California on Dec. 31?](#) market.

In the News

[How the U.S. Government Can Learn to See the Future](#) argues that rigorous probabilistic forecasting, keeping score of assessments, and employing the “wisdom of crowds” would lead to better US intelligence assessments. They also point out that forecasting projects did not survive the “valley of death”—the space between being a pilot program and being an established product for the Department of Defense which many initiatives fail to cross.

During the Obama years, the U.S. government initiated several quantitative geopolitical forecasting projects designed to complement traditional analysis methods. Between 2008 and 2018, the Defense Advanced Research Projects Agency (DARPA) and its intelligence community counterpart, the Intelligence Advanced Research Projects Activity (IARPA), launched a portfolio of a dozen prediction and forecasting initiatives. Some were very successful, such as the Aggregative Contingent Estimation (ACE) Program, which [illustrated](#) the potential utility of open-sourced forecasts by crowds. Another program [reportedly](#) had “the largest dataset on the accuracy of analytic judgments in the history of the intelligence community,” and yet another [was credited with predicting](#) the 2013 Brazilian Spring and 2014 protests in Venezuela.

Despite promising results, nearly all of these programs ended during the Trump years. Some initiatives came to a natural end; others lost bureaucratic support. The U.S. government often faces challenges when seeking to transition promising research and development efforts into programs of record. In fact, there is a foreboding name given to this trend in the defense procurement world—the “[valley of death](#).” Many U.S. government forecasting attempts failed to emerge from the valley because of failures to effectively communicate probabilities and their value to intelligence agency officials and policymakers, and bureaucratic resistance from those who feared forecasting efforts would upend their careers or the hierarchy of subject matter experts.

There has been some recent brouhaha in the [news \(archive link\)](#) about whether COVID-19 originated from a lab. The issue was previously featured in the January edition of this newsletter:

[Rootclaim](#) is a site which comes up with Bayesian calculations for public interest questions. For example, here is their page on [the source of COVID-19](#) (80%+): they start with a reasonable prior and then legibly update their initial prediction with each piece of evidence they consider. That said, their conclusion differs from that of [Metaculus](#) (<10%) and from that of casual [discussion](#) between several superforecasters on Twitter (~25%).

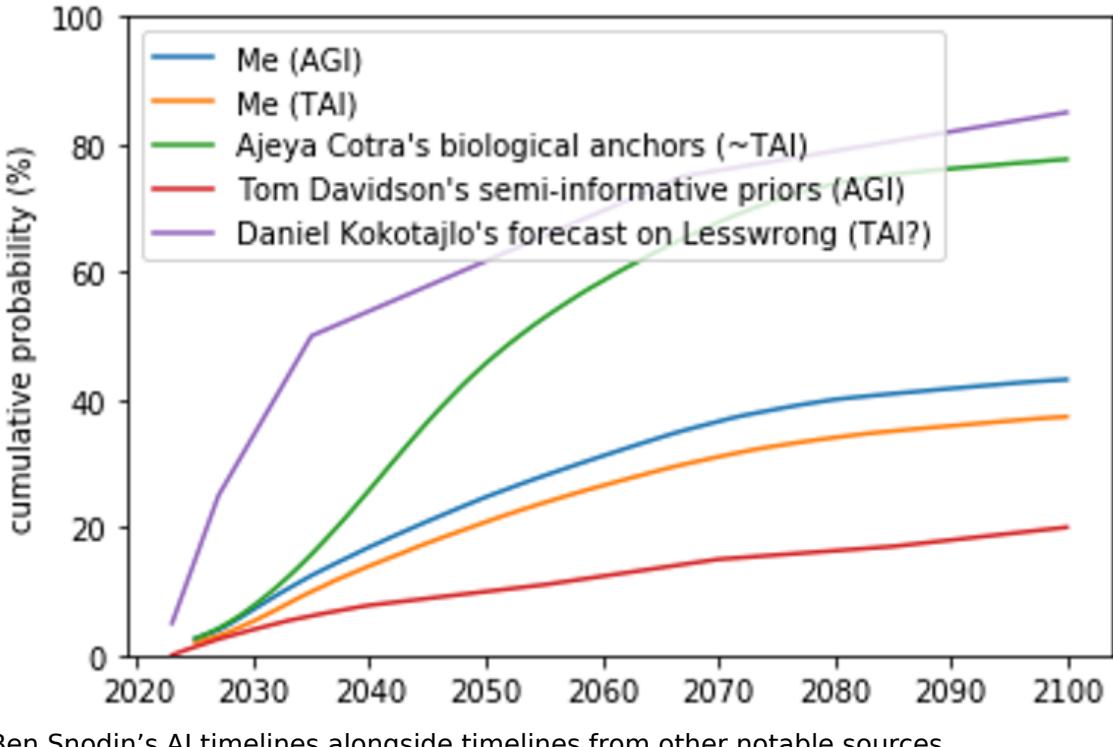
[Gauging for disasters: Neighbor shares distrust of river forecasting following flood event](#) gives a slice-of-life picture of how forecasting affects common folks. On the one hand, the interviewee is probably suffering from hindsight bias. But on the other hand, it does seem like the forecasts were not robust to further rainfall, and that grizzled grumpy locals might have had more information than the forecasters.

Blog Posts

[Probability theory does not extend logic \(predicate calculus\)](#). In particular, freely mixing logical quantifiers (\forall, \exists) and probability statements gets messy fairly quickly, and the tools to disambiguate their meaning may not be found solely in probability theory (but perhaps in statistical inference or in the study of causality.)

Probability theory can be viewed as an extension of propositional calculus. Propositional calculus is described as “a logic,” for historical reasons, but it is not what is usually meant by “logic.” Cox’s Theorem concerns only propositional calculus. Further, it was well-known long before Cox that probability theory does extend propositional calculus. Informally, probability theory can extend Aristotelian logic as well. This is usually unproblematic in practice, although it squicks logicians a bit. Probability theory by itself cannot express relationships among multiple objects, as predicate calculus (i.e. “logic”) can. The two systems are typically combined in scientific practice. In specific cases, this is intuitive and unproblematic. In general, it is difficult and an open research area.

[My attempt to think about AI timelines](#), by Ben Snodin, gives his probabilities for AI timelines based on a combination of inside and outside views, after thinking about it for 40 hours.



Ben Snodin's AI timelines alongside timelines from other notable sources

[Data on forecasting accuracy across different time horizons and levels of forecaster experience](#), by [Charles Dillon](#), builds on [earlier work by niplav](#). The post might be useful to individual forecasters seeking to learn about past failure modes when forecasting long-range questions.

We see a very well calibrated graph for predictions with <1 year time horizons, before the graph starts to sag as horizons get longer, and as with PredictionBook, things just don't seem to happen as often as predictors imagine.

[Predict responses to the "existential risk from AI" survey](#) (also on [LessWrong](#)):

I sent a short [survey](#) to ~117 people working on long-term AI issues, asking about the level of existential risk from AI; 44 responded. In ~6 days, I'm going to post the anonymized results. For now, I'm posting the methods section of my post so anyone interested can predict what the results will be.

Papers

[Decomposing the Effects of Crowd-Wisdom Aggregators: The Bias-Information-Noise \(BIN\) Model](#)

Aggregating predictions from multiple judges often yields more accurate predictions than relying on a single judge: the "wisdom-of-the-crowd" effect. This aggregation can be conducted by different methods, from simple averaging to complex techniques, like Bayesian estimators and prediction markets. This article applies a broad set of aggregation methods to subjective probability estimates from a series of geopolitical forecasting tournaments. It then uses the Bias-Information-Noise (BIN) model to disentangle three mechanisms by which each aggregation method improves accuracy: the tamping down of bias and noise and the extraction of valid information across forecasters. Averaging works almost entirely via noise reduction whereas more complex

techniques, like prediction markets and Bayesian aggregators, work via all three BIN pathways: better signal extraction and noise and bias reduction.

Hard to categorize

The Miami International Securities Exchange (MIAX) was set to offer corporate tax futures on the Minneapolis Grain Exchange (MGEX). The [site](#) currently seems to be down, but a copy remains on the [Internet Archive](#).

The United States' distopically named National Institute of Justice has a [Recidivism Forecasting Challenge](#), with a total prize pool of \$723,000, divided across many categories. I imagine that the student and small team categories should be reasonably accessible, but "individuals must be U.S. residents and companies must have an office with a U.S. business license."

The epiforecast group at the London School of Hygiene & Tropical Medicine opened the [UK Covid-19 Crowd Forecasting Challenge](#), with a prize pool of £175.

OpenPhilanthropy has a [forecasting-related job offer](#) for a relatively junior role:

Open Philanthropy is seeking a Program Assistant to support Luke Muehlhauser, who leads our work on AI policy and governance and forecasting. This role can be based out of our San Francisco office or be done fully remotely.

You will work closely with Luke to find and organize information that will accelerate and help prioritize his grantmaking, maintain and improve our internal systems for tracking and improving our forecasts, and generally help free up more of his time.

Long Content

[Imprecise probability](#) is an attempt to generalize probability theory to allow for uncertainty about or multiplicity of probability estimates. For example, consider expressing one's uncertainty by giving the odds you'd be willing to take in favor of X, and the odds you'd be willing to take against X, but those odds having a spread. See also [this summary](#) by Ben Snodin of two abstruse ivory-tower papers about the topic.

[Anthropics: different probabilities, different questions](#) dissolves the apparent paradox that different anthropic theories give different probabilities to the same event.

[The economics of faith: using an apocalyptic prophecy to elicit religious beliefs in the field](#):

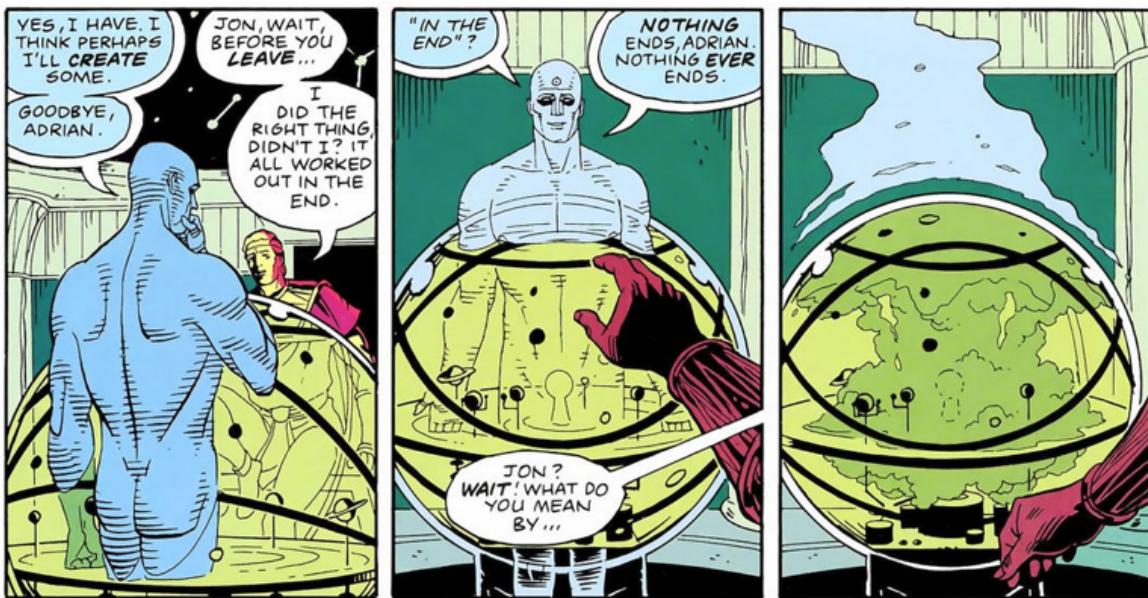
We show how standard experimental interventions linking financial consequences to falsifiable religious statements can elicit and characterize beliefs. We implemented this approach with members of a group that expected the "End of the World" to occur on May 21, 2011 by varying monetary prizes payable before and after May 21st. To our knowledge, this is the first incentivized elicitation of religious beliefs ever conducted. The results suggest that the members held extreme, sincere beliefs that were unresponsive to experimental manipulations in price.

We administered an experiment to implement our elicitation approach, relying on the well publicized prophecy made by Harold Camping, an elderly Christian radio talk show host, who held that May 21st, 2011 would be the "End of the World." On May 21st, the prophecy went, the biblical Rapture would occur: divine judgement would be passed and the "saved" would ascend to Heaven to meet God, while great cataclysms would ravage the Earth. The "non-saved" would suffer "Hell on Earth" for five months, until all of creation would be annihilated on October 21st, 2011. Camping's prediction attracted a

world-wide following, driven by tens of millions of advertising dollars and daily discussion on his Family Radio network, one of the largest Christian broadcasting networks in the U.S.

The evidence indicates that the vast majority of Family Radio members held extreme beliefs even in the face of direct financial costs—nearly all Family Radio subjects preferred \$5 dollars today to any amount up to \$500 payable after the Rapture.

Note to the future: All links are added automatically to the Internet Archive. In case of link rot, go [there](#) and input the dead link.



Alan Moore, Watchmen, [Issue #12, page 29](#)

Adrian Veidt (Ozymandias): I did the right thing, didn't I? It all worked out in the end.

Dr. Manhattan: "In the end"? Nothing ends, Adrian. Nothing ever ends.

Forecasting Newsletter: June 2021

Highlights

- Some Superforecasters start a [substack](#), as does [Dominic Cummings](#)
- Alex Lawsen and I published [Alignment Problems With Current Forecasting Platforms](#) on the arxiv.
- [What if Military AI is a Washout?](#) considers a future in which AI ends up affecting war not because of its overwhelming dominance, but by changing war's tradeoffs and best practices.

Index

- Prediction Markets & Forecasting Platforms
- In The News
- Papers
- Blog Posts

Sign up [here](#) or browse past newsletters [here](#).

Prediction Markets & Forecasting Platforms

CSET-Foretell

The Superforecasting workshop for "Foretell Pros" has ended. A tidbit I learnt from it is that, unlike prediction markets, forecasting platforms can look at the covariance between forecasters—whether two forecasters' predictions are closer or further apart than average—and update on it. That is, if two forecasters who often disagree instead agree on a question, that is evidence that their side is correct (h/t Eva Chen).

CSET has also been collaborating with Ought, and has given "Foretell Pros" access to Ought's GPT-3 assistant capabilities. I'm unclear on how often Ought's tools will be used in practice by forecasters.

Good Judgment Open

[The World Ahead: What If? \(a\)](#)—a new Good Judgment Open tournament in collaboration with The Economist—presents five long-term questions. This is something I don't recall seeing before, and I'm glad to see that Good Judgment Open is dipping its toes into the tricky business of long-term predictions. The questions will not be scored, perhaps because Good Judgment Open uses an improper scoring rule which gets worse for longer term questions, see below.

Good Judgment Open dips its toes into the tricky business of long-term predictions, and presents five questions in a new tournament [The World Ahead: What If? \(a\)](#), in collaboration with The Economist.

Metaculus

Metaculus has a new [redesign](#) (a) in progress, and an accompanying [blogpost](#) (a) by Metaculus' CEO. Some discussion can be found on Metaculus itself [here](#) (a).

Metaculus also launched the [Trade Signal tournament](#) (a), where Metaculus users attempt to predict economic indicators which might be used to make trades. For this, they are looking for a "[Community Trader](#)" (a). So far, [the one candidate](#) (a) seems very formidable.

Michael Aird, of Rethink Priorities, organized the [Nuclear Risk Forecasting Tournament](#) (a). Questions can be found [here](#) (a).

The [20/20 Insight Forecasting Contest](#) (a) has concluded. Winners can be seen [here](#) (a).

SimonM kindly curated the top comments from Metaculus this past June. They are:

- [Koji](#) writes at length about whether or not the Iran Nuclear deal will be restarted in 2021
- SimonM models potential meat demand decreasing, by reusing historical data from the "decline" in smoking
- [chrisjbillington](#) thinks the Metaculus community forecast hasn't adjusted enough for the delta variant
- [SimonM](#) models the Senate as a random walk to estimate the odds the GOP hold it for the next 10 years
- elifland_ought and EvanHarper share some forecasts for how [Japan will perform at their home Olympics](#)
- [NunoSempere](#) brings together a collection of other forecasts to estimate both the likelihood of the lab leak hypothesis AND whether or not the government will acknowledge it.
- [whaffner] (<https://www.metaculus.com/questions/7330/community-trader-election/#comment-64119>) runs for the Community Trader position
- [Trey Goff, Chief of Staff at Próspera](#), weighs in on the forecasting question about Próspera

Polymarket

Polymarket has at times been nigh-unusable because of network congestion and dependency failures. Polygon, the second layer solution for Ethereum which Polymarket uses, has been becoming more popular, so costs to make transactions (gas costs) have increased, and the infrastructure needed to process those transactions has at times been taxed beyond capacity. In response, Polymarket has increased the gas prices which its contracts were willing to pay; this doesn't really affect users because even comparatively high gas prices on Polygon are at most cents.

In addition, [The Graph](#) (a)—a service which Polymarket was relying on to let its webpage know what its blockchain contracts were doing—has also been suffering from constant failures, presumably also as a result of scaling pains.

Polygon itself has also been accused of being insecure, because 5 out 8 developers and early community members ("multisignature key holders") could conspire to

upgrade Polygon's protocol. Here are [two](#) ([a](#)) [letters](#) ([a](#)) from "DeFi Watch".

Polymarket contentiously resolved its "[Will NYC fully reopen by July 1?](#)" ([a](#)) positively. [Here](#) ([a](#)) is someone on twitter making the case for a "No" resolution, and [here](#) ([a](#)) is the case for a "Yes" resolution, whereas [here](#) ([a](#)) is Polymarket's rationale for resolving it positively, as they did. Polymarket also prematurely resolved [Will Joe Biden be President of the USA on June 30, 2021?](#) ([a](#)) as a "Yes", and they are reimbursing market participants who held "No" positions.

On the positive side, Polymarket passed its [one year anniversary](#) ([a](#)) this month, and organized a party. Some community members were invited and reimbursed for their travel expenses.

A Polymarket community member has released a [polymarket trading tool](#) ([a](#)), which allows users to interact with Polymarket's Polygon contracts directly, without having to use Polymarket's frontend. Polymarket has also added some rudimentary search functionality to their frontpage.

In more detail: Why and how could Polygon multisignature key holders steal user's funds?

A key point of contention is whether upgrading Polygon's protocol could be used to straight-out steal user's assets (or just make the platform unusable). Answering that question would require understanding some of the finer points on cross-chain communication, which are a bit beyond me. In particular, what the multisignature key holders would be stealing wouldn't directly be the valuable USDC, or ETH assets, but rather a [Doppelgänger](#) ([a](#)) of those assets, a clone asset on the Polygon Chain which is guaranteed to be redeemable for original tokens, originals which are safely stashed away in the Ethereum Chain. See: [Moving assets to Polygon](#) ([a](#)), and [wrapped tokens](#) ([a](#)).

It's possible that stealing the Doppelgänger tokens would just make them instantly worthless. More specifically, because USDC is controlled by a [central authority](#) ([a](#)), it could just refuse to honor stolen tokens. However, the malicious multisignature key holders could steal users' assets, and then very quickly swap those assets for decentralized assets (like [DAI](#)), using [Uniswap](#) ([a](#)); they could then disappear using [Tornado Cash](#) ([a](#)). This would normally not be possible, but in this case, the process to upgrade Polygon's protocol is not under a [timelock](#) ([a](#)): there is no enforced waiting period between the announcement of an upgrade and when that upgrade takes effect.

In the short term, I'm not actually too worried, and I'm keeping my assets on Polymarket, on Polygon. But in the medium to long term, the probability of things like regulatory attacks or plain old human unreliability or malice start to add up.

Superforecasters

A [shrewdness](#) ([a](#)) of superforecasters has started a [substack](#) ([a](#)), so far featuring fortnightly forecasts of [in fashion affairs](#) ([a](#)).

Others

- I added [Rootclaim](#) ([a](#)) to [Metaforecast](#) ([a](#)), and fixed a bug due to which some CSET-Foretell questions were not getting included (h/t Michał Dubrawski). I've

also rewritten the back-end code to make keeping a history of predictions feasible. This might produce some interesting comparative research in the coming months.

- [Kalshi](#) (a) may have started to allow trades, but I can't verify this because its only open to US residents.

In the News

[NPR](#) (a) has gotten some economists—who disagree with each other—to make quantifiable predictions, and to promise to come back in a couple of months to analyze what they got right or wrong: h/t @CrunchWrapSupreme.

OK, so here's what's going to happen on today's show. We're going to have two economic forecasters, Diane and Alfredo, who are going to make specific predictions about what is going to happen with jobs, with inflation and with housing in the United States for the rest of the year. Also, unlike some economic forecasters who make their predictions and then sort of disappear if they get things wrong, Alfredo and Diane have courageously agreed to come back on the show in January and talk about both what they got right and what they got wrong. And when they come back, we'll see whose forecasts were closer to reality.

[The Rise Fund Announces \\$100 Million Strategic Investment in Climavision](#) (a). The Rise Fund is one of the largest, if not the largest, impact investment funds. The investment is supposed to improve weather forecasting. Taken directly from the press release:

Climavision was formed out of Enterprise Electronics Corporation (EEC), the world's largest privately held commercial supplier of weather radar systems. EEC, which is majority controlled by the Cookes family, has [...] more than 1,200 installations across 95 countries. By combining lower altitude, proprietary data with [...] machine learning and AI technology, Climavision [...] provides [...] higher resolution and more accurate forecasting to address [...] coverage gaps left by existing radar networks across the U.S.

"As weather patterns become increasingly unpredictable and volatile due to climate change, the need for higher-quality regional and hyper local weather data has never been more pronounced," said Climavision Co-Founder and CEO Chris Goode. "Climavision's increased coverage and improved weather information enables earlier and more accurate weather forecasts that can save lives, limit business disruption, and improve the lives of people and communities across the country."

There has recently been a heat wave in the US. Compare coverage from [Fox](#) (a), from the [Associated Press](#) (a) and from [Reuters](#) (a).

[European data monopoly hurt forecasts of deadly eruption, Congolese researchers charge](#) (a).

On 22 May, Mount Nyiragongo, perhaps the most dangerous volcano in the world, erupted in a show of fire. Lava swept toward the city of Goma in the Democratic Republic of the Congo (DRC), pushing thousands from their homes and killing dozens. Although the volcano has since settled down, a new flashpoint has erupted at the geophysical observatory that monitors it.

In a 2 June open letter addressed to the DRC's president, staff at the Goma Volcano Observatory (GVO) have condemned what they say is corruption by the observatory's Congolese leadership. They also accuse European partners of a "neocolonial" attitude and of depriving them of timely data that might have allowed them to provide early warnings of eruptions.

Signed by union leader Zirirane Bijandwa Innocent on behalf of the dozens of staff researchers and technicians, the letter alleges that GVO leaders squandered money from international donors, failed to pay staff for months, and even had some researchers arrested for complaining about the situation. It also charges that the Royal Museum for Central Africa (MRAC) in Belgium and the European Centre for Geodynamics and Seismology (ECGS) in Luxembourg, long-term partners with GVO, wield too much influence over its leadership. The letter says the observatory "was taken hostage... by a small group of scientific neo-colonialists" who shut out local experts and focused on their own volcanology research at the expense of developing local capacity to monitor geohazards.

The cuts left the observatory unable to afford even an internet connection. That deprived GVO of real-time data from a network of seismometers and GPS stations deployed across the region by MRAC and ECGS since 2012. These devices can detect the small tremors and movements of Earth's surface that can precede eruptions, as magma rises inside a volcano. The sensors send their data directly to ECGS before being returned to GVO.

Another dispute concerns whether the eruption could have been predicted. In presentations at GVO on 26 April and 10 May after they regained access to the data, staff seismologists highlighted tremor activity that might indicate magma rising through cracks, according to the letter and to Science's source at the observatory. They urged GVO leadership to send teams out to make field observations, but nothing happened. The complainants allege that GVO leaders deferred to advice from their European partners.

Papers

In [Alignment Problems With Current Forecasting Platforms \(a\)](#), my coauthor Alex Lawsen and I expand upon our earlier [Incentive Problems With Current Forecasting Competitions \(a\)](#). We classify current problems as more or less either reward specification problems or more or less principal-agent problems. Reward specification problems are those which incentivize forecasters to behave in ways which are not useful from the perspective of the accuracy of the broader system.

For instance, some platforms:

- incentivize people to make forecasts on lots of questions even if they have no particular information advantage,
- disincentivize forecasters to forecast even if they know the true word-from-God probability exactly,
- strongly disincentivize people from sharing information,
- etc.

With regards to principal-agent problems, forecasters also sometimes stop trying to maximize their expected score, and instead start optimizing for other metrics. For example, discrete prizes create incentives to be in the top people who get prizes, or in

the top few spots where people can brag that they won a tournament. We try to analyze this effect quantitatively. We also prove that some platforms, like Good Judgment Open or CSET-Foretell, straight out use an improper scoring rule, where participants can get a better score in expectation by inputting something other than their true probability.

I thought that this was going to be a big deal, because Superforecasters are chosen from Good Judgment Open, but per Good Judgment Inc, the effect probably turns out to be small. As a tidbit from history, IARPA's ACE tournament also used an improper scoring rule, but other groups besides the Good Judgment Project thought that it would be too much of a hassle to change.

In any case, each of the alignment problems we identify can manifest itself in different ways. Forecasters can consciously follow their flawed incentives. But it is also the case that each alignment failure adds noise to the ranking of forecasters (even if the noise is random). More spookily, forecasters also interpret their scores (or the monetary reward in the case of a tournament) as feedback. So to the extent that this feedback is flawed, forecasters might implicitly learn the wrong lessons. This possibility is particularly worrisome to me because "the feeling of a 80%", or "the feeling of updating from an 80% to a 60%" is for me something fairly intuitive. Thus, it is something which I could imagine could be vulnerable to flawed training. See [Unconscious Economics \(a\)](#) for an elaboration of the point that incentives don't have to consciously be followed to affect outcomes.

Many of the problems above are solved by prediction markets. But prediction markets have their own [problems \(a\)](#) and [inefficiencies \(a\)](#). For example, prediction markets also greatly disincentivize collaboration and thus greatly incentivize redundancy in research (a.k.a. "have you ever seen good comments on PredictIt?" h/t Marc Koehler.)

We also propose solutions for these problems. My preferred solution right now is one in which:

- forecasters are rewarded in proportion to how well they and their team or their community do
- against a prior selected by the forecasting platform,
- the winners are not revealed, and
- rewards are either continuous or probabilistic (and as a result, proper).

However, in the setup I have in mind, the forecasting platform ends up paying money proportionally to the number of forecasters (and is thus easily exploitable), or forecasters are disincentivized to bring other people in even if they would improve probabilities. Additionally, forecasters have with an incentive to "slack-off"—to wait until someone else shares their hard work and reap similar rewards as them.

The conclusion section makes some comparisons between aligning forecasting systems and aligning machine systems. They both have a chain of proxies between the original goal and what ends up being maximized. And even though the human forecasters aren't being trained or optimized, there still seems to be a comparison to be made between the [inner alignment \(a\)](#) problem for reinforcement learners and the principal/agent problem for forecasters. Similarly, reward specification seems fairly equivalent to outer alignment, though I might be missing some nuance. I'm not really sure to what extent I'm shooting from the hip here, but I suggest that alignment proposals which would apply to superhuman systems could be tested on human forecasters with the goal of making them produce useful forecasts.

Blog Posts

[Dominic Cummings \(a\)](#) has started a substack. On the one hand, he appears to have deep insight about the inner workings of Britain's political machinery. On the other hand, it's difficult to say how Machiavellian he is, what proportion of what he communicates is intended to shape public opinion in a certain way, or how distorted his models of the world are by a goal of communicating information to have some effect. One of the things the British leave campaign did under his direction was to run randomized trials/focus groups on the most persuasive arguments for Brexit were. I remember reading them, and finding them very persuasive, and then realizing that such persuasiveness was probably fairly uncorrelated with the truth. In LessWrong lingo, I'm unsure about which [Simulacrum Level \(a\)](#) Cummings is operating at.

[Event-driven mission hedging and the 2020 US election \(a\)](#) considers a case where it is cheaper to buy some altruistic good if Biden wins, so one could bet on his success and buy it only if he wins. The post makes the mistake of ignoring market dynamics, but this doesn't change the thrust of its argument.

If Biden wins the election then, based on your research, you expect the effectiveness of your donation will rise by about 10x. Suppose the baseline cost-effectiveness of the CCF is roughly \$1/tCO₂e (tonne of CO₂ equivalent), so under Biden you expect it to be better at \$0.1/tCO₂e.

You believe Biden has a 70% chance of winning (see [NYT article \(a\)](#)). You see that on Betfair you can get 1.5:1 odds on Biden.

If you just donate, then your donation averts 7.3 million tCO₂e in expectation. See below the main post for the calculations.

But if you bet \$1m on Biden, with the commitment to donate the potential \$1.5m win, then your expected impact is to avert 10.5 million tCO₂e.

So, because almost all the impact you expect from your donation occurs when Biden wins, you can increase your expected impact by more than 40%.

[The Ultimate Guide to Decentralized Prediction Markets \(a\)](#), an old Augur blog post that covers the topic in depth.

[What to Expect When You're Expecting Inflation \(a\)](#):

In this post, I want to explore the different measures of inflation expectations in the United States and their relative accuracy in predicting actual inflation in the hopes of informing an evaluation of today's inflation expectations. I will show that inflation expectations have been well-anchored and fairly accurate, often overestimating realized inflation over the last two decades.

Jason Crawford on [precognition \(a\)](#):

Most people are slow to notice and accept change. If you can just be faster than most people at seeing what's going on, updating your model of the world, and reacting accordingly, it's almost as good as seeing the future.

[Taboo "Outside View" \(a\)](#):

The term is easily abused and its meaning has expanded too much. I recommend we permanently taboo "Outside view," i.e. stop using the word and use more precise, less confused concepts instead. This post explains why.

[The Generalized Product Rule](#) (a) outlines how a certain step in [Cox's theorem](#) (a)—the step which proves that probability updating is multiplicative—can be applied to other problems as well.

[The Perils of Forecasting](#) (a):

As mentioned earlier, the intelligence and business communities tend to be much more seasoned and thorough in their analyses of ground-breaking paradigms. That is the case because they are not involved in public grandstanding about their own cleverness to the degree that some journalists and academics are. It is also because intelligence agencies and corporations are on a mission to try to get the future right: whether for reasons of national security or the commercial profit motive. Intelligence services and businesses also know that forecasting a middle-term future of five-to-15 years is essential, and yet they are aware just how difficult it is. They know that linear thinking is hard to escape from, since extrapolating from current trends is often all one ever has to go on. So, they are understanding of attempts at non-linear analysis, even when flawed. And because corporations and businesses meet behind closed doors, they are more willing to countenance blunt, hard-nosed assessments about such things as national cultures than journalists and academics are.

Journalists, contrarily, are consumed with presentness. They tend to judge everything from the vantage point of the current news cycle. And it is this obsession with presentness that obscures historical context, from which the future can be discerned, however imperfectly.

Interestingly, at a time when even the finest elite publications do not cover foreign affairs as seriously and as disinterestedly as they once did, corporations have been reaching out to private forecasting companies to get a cold-blooded sense of the middle-term future in many places. Having worked for two such firms—Stratfor and Eurasia Group—over the course of the recent decade, I can confirm that even when wrong, what such firms really bring to the table is an old-fashioned and comprehensive seriousness about the news of the world and where it is headed, regardless of its human interest value. They also are deliberately amoral: whether an outcome is good or bad does not interest them as a firm. The point is whether they predicted it or not.

And the more that the media as a whole declines—trafficking in the trivial and remaining within predictable philosophical comfort zones—the more necessary such firms will be. Indeed, the media is dominated by liberal arts majors, who are driven by the need to turn the stories of individuals into narratives; whereas analysis—the weighing of harsh, unpleasant truths that require abstractions and generalizations—is often the pursuit of math minds.

[What if Military AI is a Washout?](#) (a). The author presents his "hunches" on the future of military AI, in which it does improve, but it ends up affecting war not because of its overwhelming dominance, but by changing the tradeoffs and best practices of war. For instance, war might move more and more into cities, because they are an environment in which classifier systems might be more uncertain about whether someone is a civilian or an enemy combatant.

Here are some hunches about how I think military AI plays out. I use hunches here because I think hunches are a better currency than predictions. Predictions are ten-a-penny and always subject to retroactive revision ("Well I said that AI was going to transform warfare but if you look at it this way then cleaning the operations room with a Roomba is transformative"). Hunches are like predictions but without the veneer of professional expertise. Everyone can have hunches. Hunches are often more descriptive of underlying thinking than they are of the end product, so to speak. Like predictions, hunches require little to no support, but in terms of plain language they are far more open about this fact.

Integrated AI systems into organisational processes distorts them where it is possible (just as integrating desktop computers, or typewriters, or new production processes changes an organisation). If you have a factory where you can present a computer model of a required output and the factory itself will optimise the tooling and production lines to make it, then you'll get a jump on competitors if you compete in terms of taking novel items to market. If, on the other hand, you make an error in the model, then model errors are now likely much more expensive, as there would be less time to identify and rectify them before producing the (expensive) tools to mass produce them. Result: less people in tooling, more people in model quality assurance. In my view, the same goes for targeting processes. If bits of the "kill chain" get automated with AI, then it increases risks of prior incorrect human (or machine) judgement. Result: re-shaping military organisations to account for potential optimisations offered by AI, and to minimise risks of errors.

In this view artificial intelligence is essentially automation. We take something that would require human cognition and action, instantiate it in a physical system, and then something that used to require a human being no longer requires a human being. "That is not AI" I hear you say, well, in response, consider how many automatic things were once autonomous things. Fire-and-forget missiles have gone from being discussed as autonomous systems to simply being an automatic function of a system. Automated Target Recognition systems are performing equivalent cognitive work (recognising objects from sense data) to human beings. It's just that they can make sense of many different types of data, and do it faster than we can, enabling forms of action beyond human capabilities.

As I see it, object recognition is a key domain in which AI will eventually outperform us. At least for big recognisable pieces of kit. Therein lies the asymmetry - big pieces of recognisable military kit will be vulnerable to recognition by autonomous systems, whereas distinguishing human beings as being combatants or civilians is going to be hard, if not impossible to achieve.

Note to the future: All links are added automatically to the Internet Archive, using [this tool](#) (a). "(a)" for archived links was inspired by [Milan Griffes](#) (a), [Andrew Zuckerman](#) (a), and [Alexey Guzey](#) (a).

Forecasting Newsletter: July 2021

Highlights

- [Biatob](#) is a new site to embed betting odds into one's writing
- [Kalshi](#), a [CFTC](#)-regulated prediction market, launches in the US
- [Malta](#) in [trouble](#) over betting and gambling fraud

Index

- Prediction Markets & Forecasting Platforms
- In The News
- Blog Posts
- Long Content

Sign up [here](#) or browse past newsletters [here](#).

Prediction Markets & Forecasting Platforms

Metaculus

[SimonM](#) (a) kindly curated the top comments from Metaculus this past July. They [are](#) (a):

- [SimonM](#) (a), [alexrlj](#) (a), [alexrlj & discussion](#) (a), [alexrlj](#) (a), and [alexrlj](#) (a) make good points on the recent COVID-19 wave in the UK.
- [Uncle Jeff](#) (a) and [JonathanShi](#) (a) don't think commercial animal farming is going anywhere.
- [NunoSempere](#) (a) and [elifland_ought](#) (a) make the case for a higher prediction of a civil war in the US.
- [Charles](#) (a) looks at the chances of a US-Russia war in the next 30 years
- [ege_erdil](#) (a) looks at the base rates on inadvertent nuclear detonations
- [PeterWildeford](#) (a) links to FiveThirtyEight's new [Olympic medal tracker](#) (a)
- [elifland_ought](#) (a) has some suggestions for better metrics for predictors.
- [KnowName](#) (a) explains why questions should close earlier

Round 2 of the [Keep Virginia Safe Tournament](#) (a) will begin in early August. It'll focus on Delta and other variants of concern, access to and rollout of the vaccine, and the safe reopening of schools in the fall.

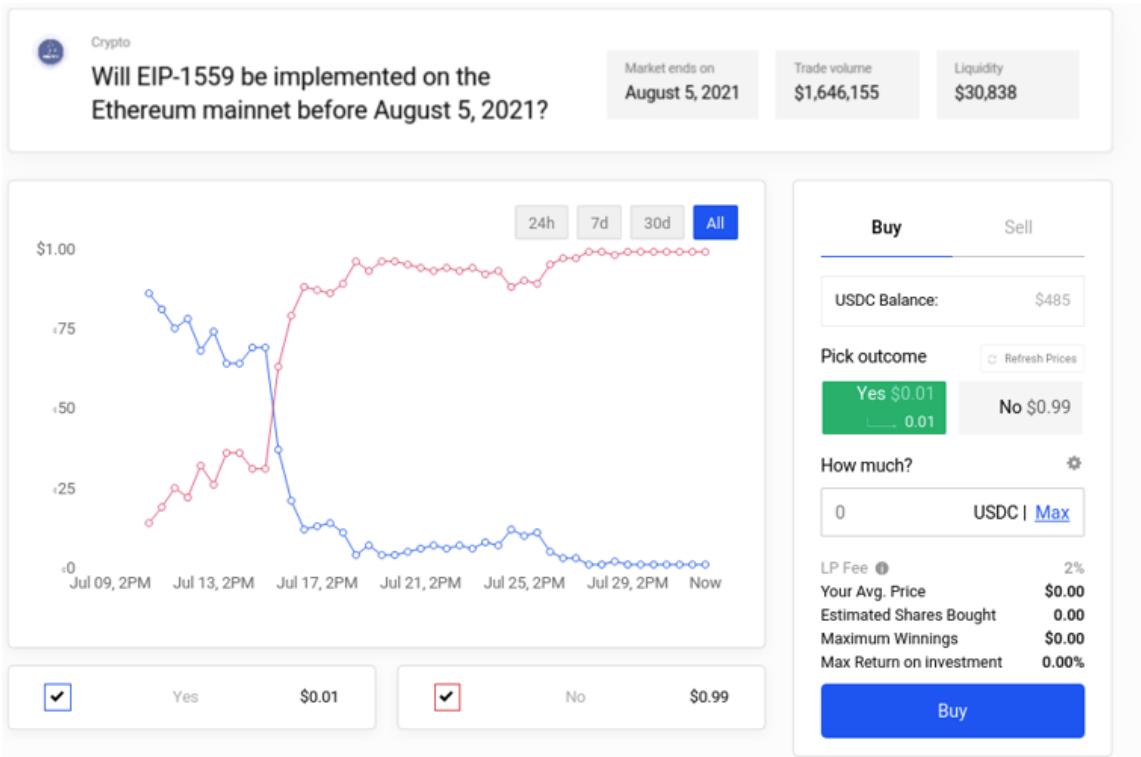
[Charles Dillon](#)—a Rethink Priorities volunteer—created a Metaculus series on [Open Philanthropy's donation volumes](#) (a). Charles also wrote [an examination of Metaculus' resolved AI predictions and their implications for AI timelines](#) (a), which tentatively finds that the Metaculus community expected slightly more progress than actually occurred.

Polymarket

Polymarket had several prominent cryptocurrency prediction markets. [Will Cardano support smart contracts on Mainnet by October 1st, 2021?](#) (a) called Cardano developers out on [missed deadlines](#) (a) ([secondary source](#) (a)).

[Will EIP-1559 be implemented on the Ethereum mainnet before August 5, 2021?](#) (a) saw Polymarket pros beat Ethereum enthusiasts by more accurately calculating block times.

Lance, an expert predictor market player, covers the topic [here](#) (a).



Polymarket also started their first tournament, the first round of which is currently ongoing. 32 participants each received \$100, and face-off in a [sudden-death tournament](#) (a). Participants' profits can be followed on [PolymarketWhales](#) (a).

Kalshi

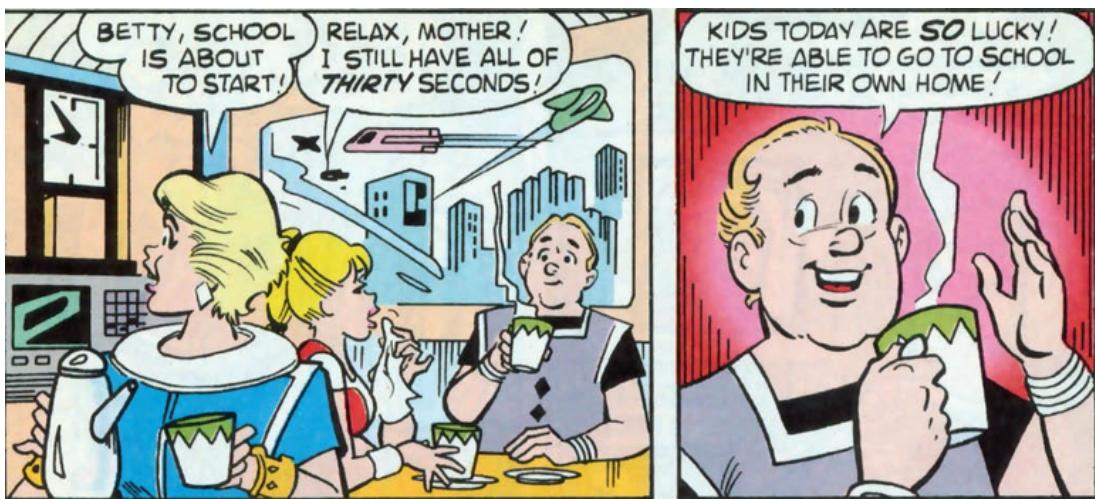
[Kalshi](#) (a)—a [CFTC](#) (a)-regulated prediction market—has launched, and is now available to US citizens. Kalshi previously raised [\\$30 million](#) (a) in a round led by [Sequoia Capital](#) (a). [Fees](#) (a) are significantly higher than those of Polymarket.

Reddit

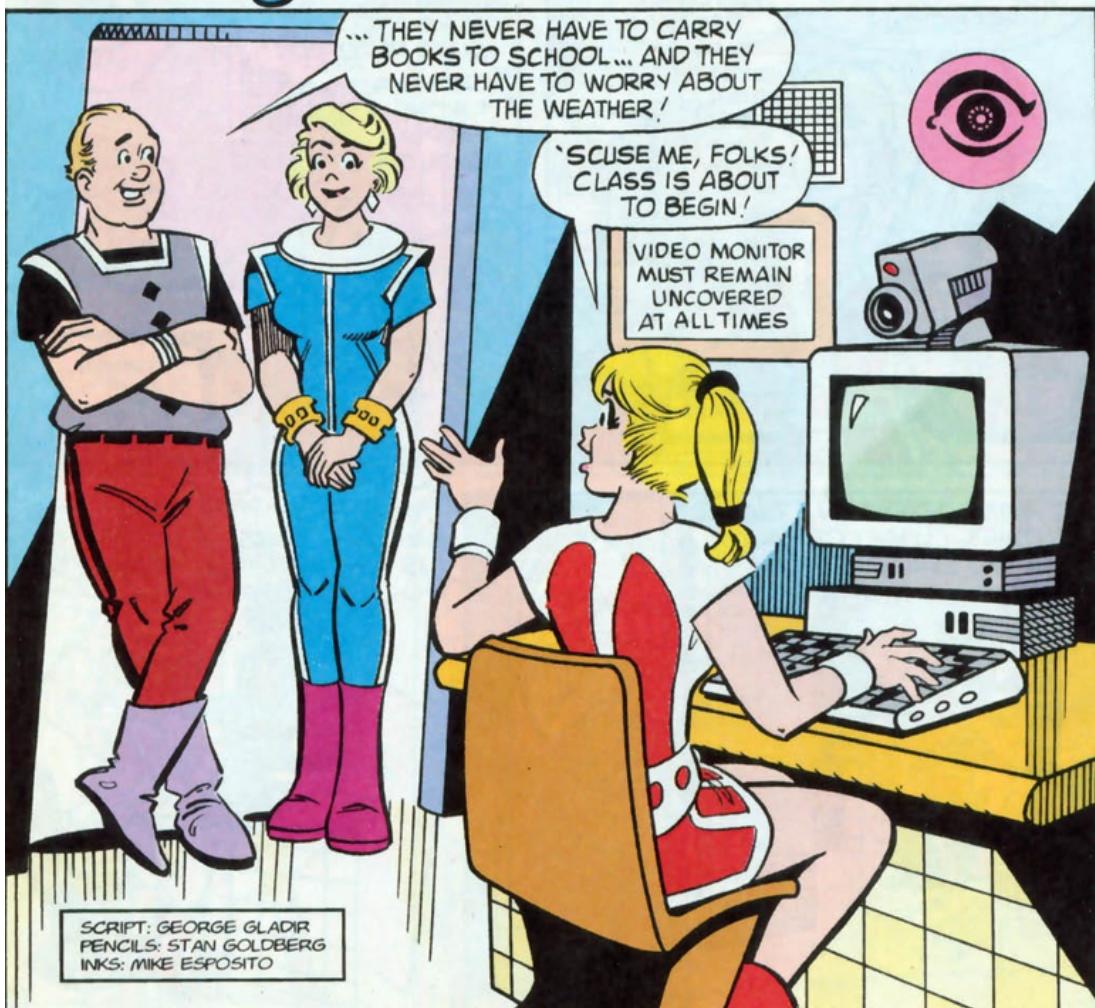
Reddit added some [prediction functionality](#) (a) late last year, and the NBA subreddit has recently been [using it](#) (a). See [Incentivizing forecasting via social media](#) (a) and [Prediction markets for internet points?](#) (a) for two posts which explore the general topic of predictions on social media platforms.

Also on Reddit, [r/MarkMyWords](#) (a) contains predictions which people want remembered, and [r/calledit](#) (a) contains those predictions which people surprisingly got right. Some highlights:

- [Betty In School in 2021 AD \(originally from Feb. 1997\)](#) (a) ([original source](#) (a))



Betty® HIGH SCHOOL 2021 A.D.



- [News anchorman predicts that Trump could win the Republican nomination in 2015, gets laughed at \(a\) \(original source \(a\)\)](#)

- [Just a reminder this guy correctly predicted the end of How I Met Your Mother.](#) ([a](#)) ([original source](#) ([a](#)))
- [Mark my words, Ruth Bader Ginsburg will die by September and Trump and the Republicans will ram through a replacement before Biden is sworn in.](#) ([a](#))

The predictions from r/MarkMyWords about public events could be tallied to obtain data on medium to long-term accuracy, and the correct predictions from r/calledit could be used to get a sense of how powerful human hypothesis generation is.

Hedgehog Markets

[Hedgehog Markets Raises \\$3.5M in Seed Funding](#) ([a](#)). They also give a [preview](#) ([a](#)) of their upcoming platform. I'm usually not a fan of announcements of an announcement, but in this case I thought it was worth mentioning:

A No-Loss Competition allows users to make predictions on event outcomes without losing their principal.

It works like this — a user decides they are interested in participating in one of Hedgehog's No-Loss Competitions. So, they stake USDC to receive game tokens, and use those game tokens to participate in various prediction markets offered within the competition. Regardless of how a user's predictions perform, they will always receive back their original USDC stake at the end of the competition.

The problem this solves is that within the DeFi ecosystem, the time value of money—the amount of interest one can earn from money by letting it sit idle, e.g., by lending it to other people or by lending liquidity to stable-coin pools—is fairly high. So once one is willing to get one's money into a blockchain, it's not clear that betting offers the best return on investment. But with Hedgehog Market's proposed functionality, one can get the returns on betting plus the interest rate of money at the same time.

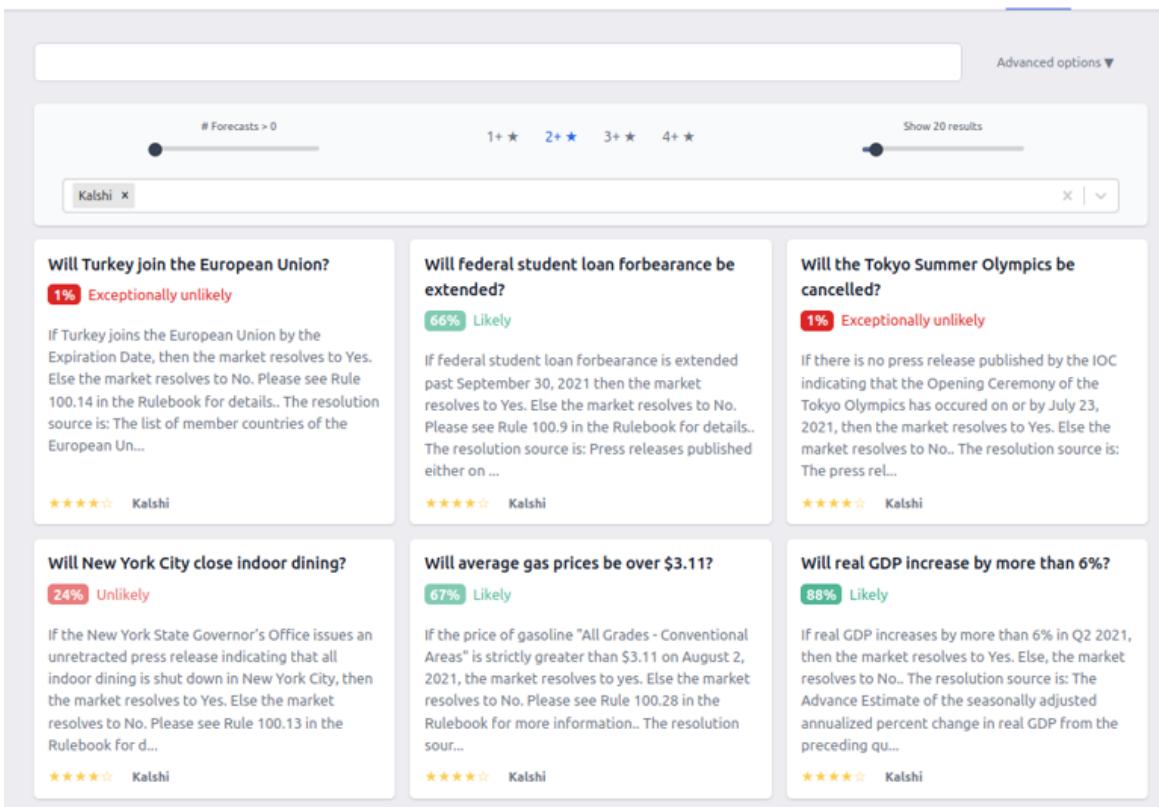
In practice, the proposed design isn't quite right, because Hedgehog contests unnecessarily consist of more than one question, and because one can't also bet the principal. But in the long run, this proposal, or others like it, should make Polymarket worried that it will lose its #1 spot as the best crypto prediction market.

Odds and Ends

[Biatob](#) ([a](#))—an acronym of "Betting is a Tax On Bullshit"—is a new site for embedding betting odds on one's writing. Like: I [\(bet: \\$20 at 50%\)](#) ([a](#)) that this newsletter will exceed 500 subscribers by the end of 2021. [Here](#) ([a](#)) is a LessWrong post introducing it.

Hypermind launches a new contest on [the future of AI](#) ([a](#)), with 30,000€ at stake for prizes. An interview with Jacob Steinhardt, a UC Berkeley professor who selected the questions, can be found [here](#) ([a](#)). Hypermind's website has also undergone a light redesign.

I've added [Kalshi](#) and [Betfair](#) to [Metaforecast](#).



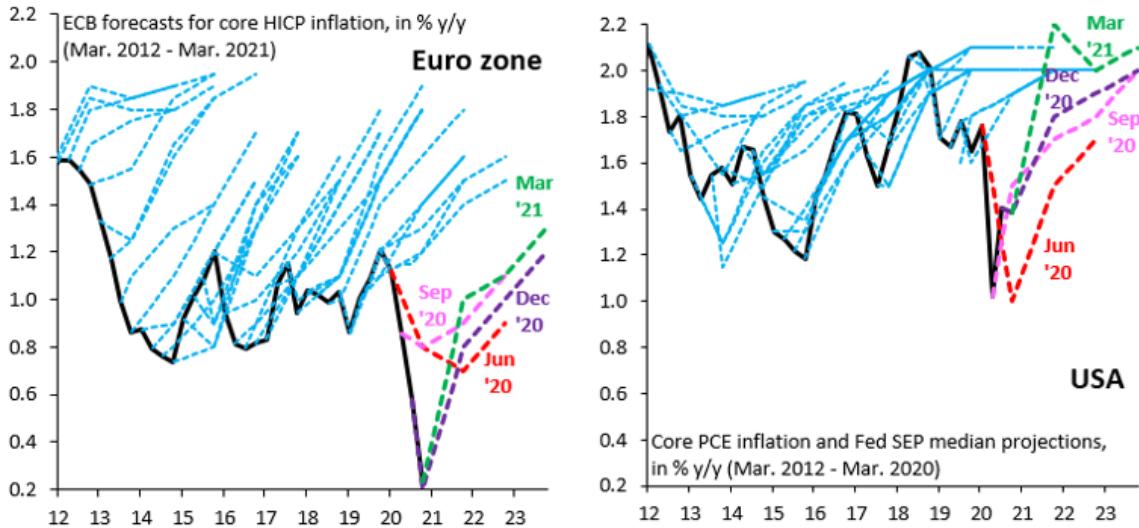
In the News

[Unfortunately, Fabs Won't Fix Forecasting \(a\)](#) gives a brief overview of the state of the semiconductor manufacturing industry. The recent chips shortage has led to more fabs being built to serve anticipated demand, and to tighter coordination between buyers and manufacturers. The article then makes a point that "...companies are looking for ways to mitigate shortages. Building fabs is part of the answer, but unless OEMs ([original equipment manufacturers \(a\)](#)) and the supply chain can improve the accuracy of their forecasts, the chip industry's next problem could be overcapacity."

Malta is in trouble over betting & fraud: [Malta faces EU sports betting veto withdrawal \(a\)](#) & [Malta first EU state placed on international money laundering watch-list \(a\)](#). H/t Roman Hagelstein. From the first article:

As one of Europe's most prominent gambling hubs – online gambling accounts for 12% of the island's GDP, generating €700 million and employing 9,000 people – and providing a base to over 250 betting operators including Betsson, Tipico and William Hill, the new stipulations could have a substantial impact on the day-to-day functions of Malta's economy.

The European Central Bank seems to systematically [over-predict inflation \(a\)](#).



Forecasting Swine Disease Outbreaks (a)

For many years, production companies have been reporting the infection status of their sow farms to the MSHMP. So now we have this incredible dataset showing whether any given farm is infected with porcine epidemic diarrhea (PED) virus in a given week. We combine these data with animal movement data, both into the sow farms as well as into neighboring farms, to build a predictive, machine-learning algorithm that actually forecasts when and where we expect there to be high probability of a PED outbreak.

The forecasting pipeline has a sensitivity of around 20%, which means that researchers can detect one out of every five outbreaks that occur.

That's more information than we had before... so it's a modest improvement," VanderWaal said. "However, if we try to improve the sensitivity, we basically create more false alarms. The positive predictive value is 70%, which means that for every 10 times the model predicts an outbreak, it's right seven of those times. Our partners don't want to get a bunch of false alarms; if you 'cry wolf' too often, people stop responding. That's one of the limitations we're trying to balance.

Blog Posts

[Thinking fast, slow, and not at all: System 3 jumps the shark \(a\)](#): Andrew Gelman tears into Kahneman's new book *Noise*; Kahneman answers in the comments.

Something similar seems to have happened with Kahneman. His first book was all about his own research, which in turn was full of skepticism for simple models of human cognition and decision making. But he left it all on the table in that book, so now he's writing about other people's work, which requires trusting in his coauthors. I think some of that trust was misplaced.

Superforecasters look at the chances of [a war over Taiwan \(a\)](#) and at [how long Kabul has left after America's withdrawal from Afghanistan \(a\)](#).

In [Shallow evaluations of longtermist organizations \(a\)](#), I look at the pathways to impact for a number of prominent longtermist EA organizations, and I give some quantified estimates of their impact or of proxies of impact.

Global Guessing [interviews Juan Cambeiro](#) (a)—a superforecaster known for his prescient COVID-19 predictions—and goes over three forecasting questions with him. Forecasters who are just starting out might find the description of what steps Juan takes when making a forecast particularly valuable.

[Types of specification problems in forecasting](#) (a) categorizes said problems and suggests solutions. It's part of a broader set of forecasting-related posts by [Rethink Priorities](#) (a).

[Risk Premiums vs Prediction Markets](#) (a) explains how risk premiums might distort market forecasts. For example, if money is worth less when markets are doing well, and more when markets are doing worse, a fair 50:50 bet on a 50% outcome might have negative expected utility. The post is slightly technical.

[Leaving the casino](#) (a). "Probabilistic rationality was originally invented to choose optimal strategies in betting games. It's perfect for that—and less perfect for other things."

[16 types of useful predictions](#) (a) is an old LessWrong post by Julia Galef, with some interesting discussion in the comments about how one can seem more or less accurate when comparing oneself to other people, depending on the method of comparison.

Long Content

[The Complexity of Agreement](#) (a) is a classical paper by Scott Aaronson which shows that the results of [Aumann's agreement theorem](#) hold in practice.

A celebrated 1976 theorem of Aumann asserts that Bayesian agents with common priors can never “agree to disagree”: if their opinions about any topic are common knowledge, then those opinions must be equal. But two key questions went unaddressed: first, can the agents reach agreement after a conversation of reasonable length? Second, can the computations needed for that conversation be performed efficiently? This paper answers both questions in the affirmative, thereby strengthening Aumann’s original conclusion.

We show that for two agents with a common prior to agree within ϵ about the expectation of a $[0, 1]$ variable with high probability over their prior, it suffices for them to exchange $O(1/\epsilon^2)$ bits. This bound is completely independent of the number of bits n of relevant knowledge that the agents have. We also extend the bound to three or more agents; and we give an example where the “standard protocol” (which consists of repeatedly announcing one’s current expectation) nearly saturates the bound, while a new “attenuated protocol” does better

This paper initiates the study of the communication complexity and computational complexity of agreement protocols. Its surprising conclusion is that, in spite of the above arguments, complexity is not a fundamental barrier to agreement. In our view, this conclusion closes a major gap between Aumann’s theorem and its informal interpretation, by showing that agreeing to disagree is problematic not merely “in the limit” of common knowledge, but even for agents subject to realistic constraints on communication and computation

In Section 4 we shift attention to the computational complexity of agreement, the subject of our deepest technical result. What we want to show is that, even if two agents are computationally bounded, after a conversation of reasonable length they can still probably approximately agree about the expectation of a $[0, 1]$ random variable. A large part of the problem is to say what this even means. After all, if the agents both ignored their evidence and estimated (say) $1/2$, then they would agree before exchanging even a single message. So agreement is only interesting if the agents have made some sort of “good-faith effort” to emulate Bayesian rationality.

The blog post [The Principle of Indifference & Bertrand's Paradox](#) (a) gives very clear examples of the problem of priors. It's a chapter from a [free online textbook](#) (a) on probability.

What's the problem? Imagine a factory makes square pieces of paper, whose sides always have length somewhere between 1 and 3 feet. What is the probability the sides of the next piece of paper they manufacture will be between 1 and 2 feet long?

Applying the Principle of Indifference we get 1/2

That seems reasonable, but now suppose we rephrase the question. What is the probability that the area of the next piece of paper will be between 1 ft² and 4 ft²? Applying the Principle of Indifference again, we get a different number, 3/8

But the answer should have been the same as before: it's the same question, just rephrased! If the sides are between 1 and 2 feet long, that's the same as the area being between 1 ft² and 4 ft².

The infamous [Literary Digest poll of 1936](#) (a) predicted that Roosevelt's rival would be the overwhelming winner. After Roosevelt instead overwhelmingly won, the magazine soon folded. Now, a new analysis [finds that](#) (a):

If information collected by the poll about votes cast in 1932 had been used to weight the results, the poll would have predicted a majority of electoral votes for Roosevelt in 1936, and thus would have correctly predicted the winner of the election. We explore alternative weighting methods for the 1936 poll and the models that support them. While weighting would have resulted in Roosevelt being projected as the winner, the bias in the estimates is still very large. We discuss implications of these results for today's low-response-rate surveys and how the accuracy of the modeling might be reflected better than current practice.

[Proebsting's paradox](#) (a) is an argument that appears to show that the Kelly criterion can lead to ruin. Its resolution requires understanding that "Kelly's criterion is to maximise expected rate of growth; only under restricted conditions does it correspond to maximising the log. One easy way to dismiss the paradox is to note that Kelly assumes that probabilities do not change."

Note to the future: All links are added automatically to the Internet Archive, using [this tool](#) (a). "(a)" for archived links was inspired by [Milan Griffes](#) (a), [Andrew Zuckerman](#) (a), and [Alexey Guzey](#) (a).

Forecasting Newsletter: August 2021

Highlights

- Despite forecaster consensus to the contrary, Kabul has fallen
- CSET-Foretell attempts to [make inroads](#) into US decision-making mechanisms
- The US' CDC to [launch a new outbreak analysis and forecast center](#)

Index

- Prediction Markets & Forecasting Platforms
- Blog Posts
- Long Content
- In The News

Sign up [here](#) or browse past newsletters [here](#).

Prediction Markets & Forecasting Platforms

Metaculus

[SimonM](#) (a) kindly curated the [top comments from Metaculus this past August](#) (a). They are:

- [johnnycaffeine](#) (a) asks if the community prediction is a good indicator that people didn't foresee the rapidity of the Taliban takeover
- [j.m.](#) (a) did the (now moot) impeachment math for Cuomo: there would probably have been enough votes for an impeachment to go through.
- [Jgalt](#) (a) flags up a rather poorly aged forecast by Biden.
- [alexjl](#) (a) points out a design flaw in [Rootclaim's Challenge](#) (a): "I find it ironic in the extreme that Rootclaim makes repeated reference to the overconfidence of experts, but that their challenge requires you to "win a debate", meaning that if you think they are overconfident but not directionally wrong (e.g. assigning 90% to something which you think should be assigned a 60% probability) there is no way for you to win the bet."
- [j.m.](#) (a) points out the internet was adopted slower than one might otherwise think.

Metaculus begins the next round of the [Keep Virginia Safe Tournament](#) (a), which has a \$1,000 prize pool. They have also pushed a redesign of their [frontpage](#) (a).

Good Judgment & Good Judgment Open

Here are the top few best comments from [Good Judgment Open](#) (a), as curated by myself:

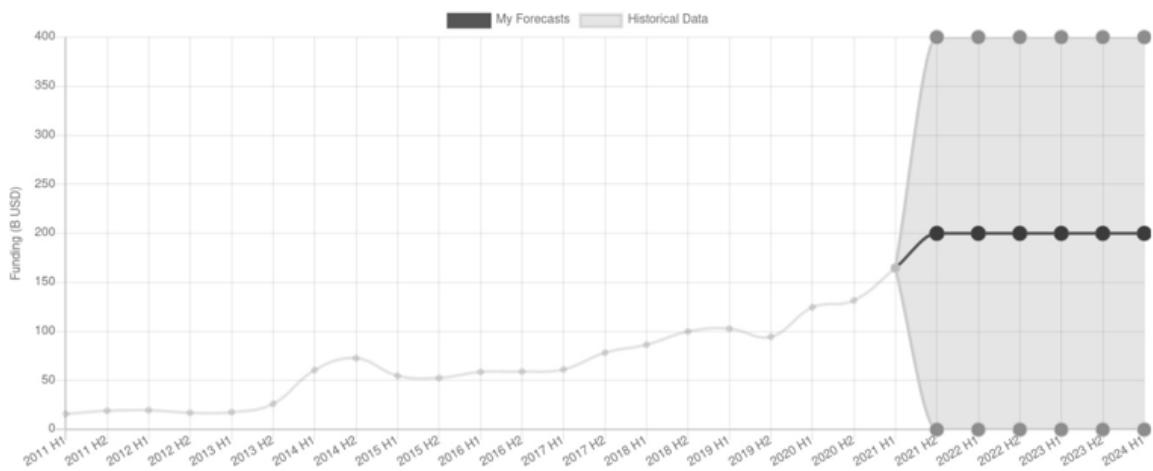
- [rjfmgv](#) (a) compares and contrasts the current situation in Afghanistan with the Bashar al-Assad regime in Syria in 2015, on which he also forecasted at the time.
- [Anneinak](#) (a) meets with the Dutch Ambassador to forecast "When will a new Dutch government be sworn in after the 2021 general election?"
- [RyanBeck](#) (a) reads Merck's Q2 earnings call document, and concludes that molnupiravir will most likely not be approved by the US FDA for use to treat COVID-19 before 1 October 2021. He also shares [this report](#) (a) by the US's Office of the Director of National Intelligence on the origins of COVID-19.
- [keith-huggins](#) (a) gives the lowdown on the power grab by the President of Tunisia.

In addition, Good Judgment Open introduced the [Sky News Challenge](#) (a). Questions are very UK-centric. But on the other hand, Sky News is a well-regarded major [UK news channel](#) (a) with a [large viewership](#) (a).

Sky News is asking forecasters for their latest judgments on political questions of consequence in the UK and beyond. Our ambition is to use probability changes in crowd forecasts to visualize how key issues in the news agenda are developing. Forecasts and reasoning that attract significant upvotes could be featured in our reporting.

CSET-Foretell

Foretell continues their work of making forecasts and forecasting methodologies more accessible and legible to US decision-makers. Most recently, they are doing this through their "Issue campaigns"; a write-up of the idea can be seen [here](#) (a). One of their first issue campaigns is on the topic of [the future of the Defense-Silicon Valley Relationship](#) (a).



Otherwise, CSET-Foretell (or rather, CultivateLabs, the company behind the webpage they use) has also [added two new question formats](#) (a). The first asks about 80% confidence intervals for more than one time period at a time, aggregating multiple sub-questions into one. The second format consists of rolling predictions, such that, e.g. "in the next six months" resets every month to refer to the next six months from the month of the forecast.

For example, a forecast on the question [Will the Chinese military or other maritime security forces fire upon another country's civil or military vessel in the South China Sea in the next six months?](#) (a) made during the month of September refers to the September-February period, whereas a forecast made during October would refer to the October-April six month period.

Personally, while I'm glad to see experimentation with new formats, and I even intellectually admit that the new formats are in a sense superior, I still find them fairly unintuitive to forecast on.

Lastly, CSET-Foretell forecasts were quoted by [Quartz](#) (a) on [on whether VC funding for tech startups will dry up](#) (a) (warning: paywalled), and by [SupChina](#) (a) on the composition of the Politburo Standing Committee of the Chinese Communist Party.

Polymarket

Difficulties with The Graph—the service which Polymarket uses to bring the data from the blockchain to its webpage—have continued, with users seeing negative balances. Polymarket has also been hosting many sports markets recently. While they bring large amounts of volume, they make Polymarket less differentiated.

A community driven [command-line trading tool](#) (a) continues to get better.

[Star Spangled Gamblers](#) (a), a political betting podcast, has been producing some highly entertaining Polymarket related podcasts, e.g.: [Betting to #FreeBritney](#) (a) and [Your California Recall Playbook is Here](#) (a).

An unrelated project in the Matic chain with a similar sounding name, [Poly Network](#) (a), was hacked for \$611 million, though most of the funds were [later returned](#) (a). Polymarket users pretended to be confused about this out of that perverse sense of humor prevalent in our time.

Metaforecast

I've pushed some [major upgrades](#) (a) to [Metaforecast](#) (a). Chiefly:

- search is much better,
- one can capture forecasts as images, making it easier to incorporate into blogposts,
- questions have quality indicators (number of forecasters, volume traded, liquidity, etc.),
- and I've added new platforms

In addition, I've written a Twitter bot which answers with the closest prediction in the database when @metaforecast is [mentioned in a tweet](#) (a):



Nuño Sempere @NunoSempere · Aug 24

Nuclear risk @metaforecast

...

2



3



Metaforecast

@metaforecast

...

Replies to @NunoSempere

Will a non-test nuclear detonation by a non-state actor cause at least one fatality by 2030? metaculus.com/questions/7406...

See also: metaforecast.org/?query=Nuclear...

Will a non-test nuclear detonation by a non-state actor cause at least one fatality by 2030?

10% Unlikely

Last updated: 2021-08-19

A nuclear detonation by a non-state actor could potentially have dire consequences either directly or via triggering other harmful actions by other actors. Non-state actors have never yet detonated nuclear weapons, but there are conceivable scenarios...



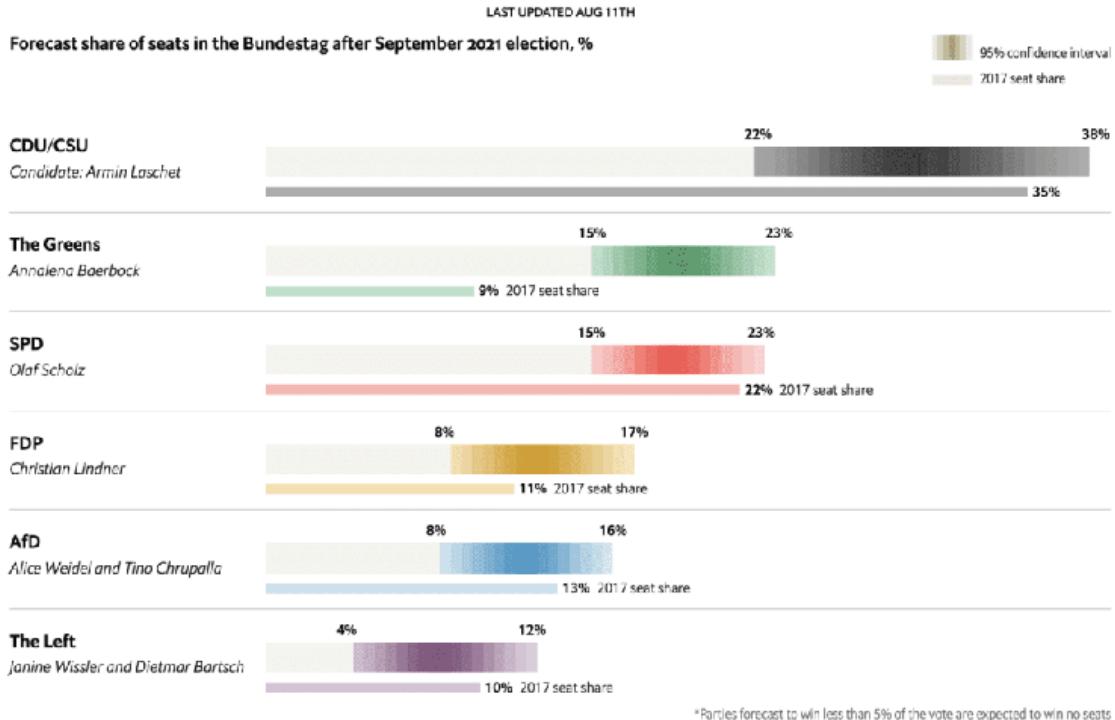
Metaculus

Forecasts: 29

This makes mentioning forecasts in casual twitter conversation pretty much trivial, so perhaps the sanity level of Twitter conversations could be raised ever so slightly.

Blog Posts

[Statistical Modelling](#) (a) discusses [Forecast displays that emphasize uncertainty](#) (a), on account of [The Economist's Forecasting Model for the 2021 German election](#) (a) ([unpaywalled version](#)), which only offers 95% confidence intervals of parliament seats, rather than explicit probabilities of who will win.



[Uncertainty can Defuse Logical Explosions](#) (a) makes the point that the [principle of explosion](#) (a)—"P and $\neg P$, therefore anything follows"—does not apply for an agent with probabilistic beliefs. I thought that the point was very neat, but also that it could be formalized better.

Daniel Kokotajlo writes [What 2026 looks like \(Daniel's Median Future\)](#) (a), extrapolating the performance of models like GPT-3 year by year.

The team behind [Global Guessing](#) (a) is starting a monthly newsletter focused on prediction markets: [Crowd Money](#) (a).

Long Content

[The D-Squared Digest One Minute MBA—Avoiding Projects Pursued By Morons 101](#) (a). A blogger which correctly predicted that there would be no weapons of mass destruction in Iraq looks back at how and why.

Literally people have been asking me: "How is it that you were so amazingly prescient about Iraq? Why is it that you were right about everything at precisely the same moment when we were wrong?" No honestly, they have. I'd love to show you the emails I've received, there were dozens of them, honest. Honest. Anyway, I note that "errors of prewar planning" is now pretty much a mainstream stylised fact, so I suspect that it might make some small contribution to the commonweal if I were to explain how it was that I was able to spot so early that this dog wasn't going to hunt. I will struggle manfully with the savage burden of boasting, self-aggrandisement and ego-stroking that this will necessarily involve. It's been done before, although admittedly by a madman in the process of dying of syphilis of the brain. Sorry, where was I?

In the [Abilene paradox](#) (a), a group of people collectively decide on a course of action that is counter to the preferences of many or all of the individuals in the group. It involves a common breakdown of group communication in which each member mistakenly believes that their own preferences are counter to the group's and, therefore, does not raise

objections. A common phrase relating to the Abilene paradox is a desire to not "rock the boat". h/t Chana.

In the News

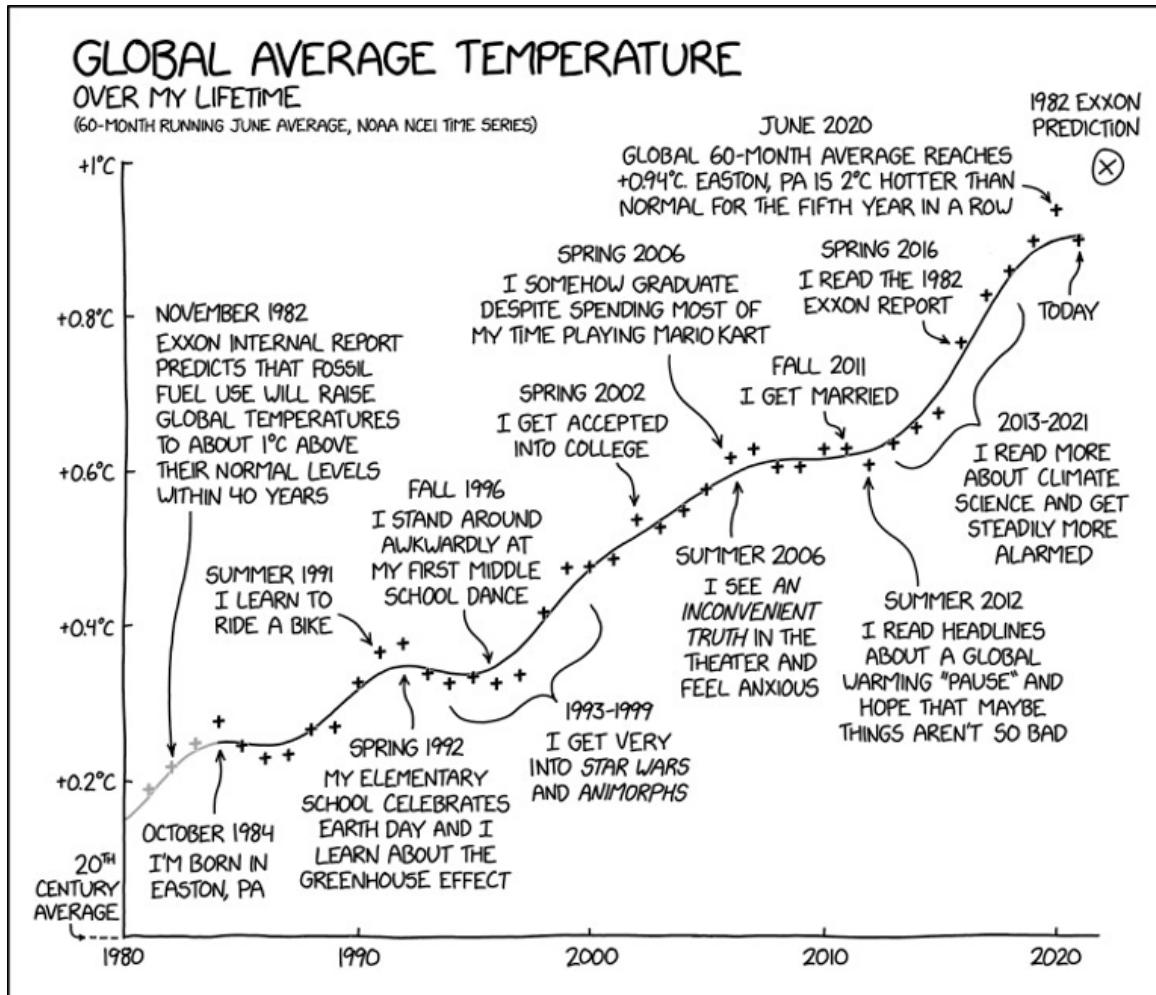
The United Nations' [Intergovernmental Panel on Climate Change \(a\)](#) has a new [report \(a\)](#) out. The report has been making the rounds; e.g., Boris Johnson described it as "[sobering reading](#)" ([a](#)). Interestingly, it uses probabilistic quantifiers:

Each finding is grounded in an evaluation of underlying evidence and agreement. A level of confidence is expressed using five qualifiers: very low, low, medium, high and very high, and typeset in italics, for example, medium confidence. The following terms have been used to indicate the assessed likelihood of an outcome or a result: virtually certain 99–100% probability, very likely 90–100%, likely 66–100%, about as likely as not 33–66%, unlikely 0–33%, very unlikely 0–10%, exceptionally unlikely 0–1%. Additional terms (extremely likely 95–100%, more likely than not >50–100%, and extremely unlikely 0–5%) may also be used when appropriate. Assessed likelihood is typeset in italics, for example, *very likely*. This is consistent with AR5. In this Report, unless stated otherwise, square brackets [x to y] are used to provide the assessed very likely range, or 90% interval

For example:

Human influence is *very likely* the main driver of the global retreat of glaciers since the 1990s and the decrease in Arctic sea ice area between 1979–1988 and 2010–2019 (about 40% in September and about 10% in March). There has been no significant trend in Antarctic sea ice area from 1979 to 2020 due to regionally opposing trends and large internal variability. Human influence *very likely* contributed to the decrease in Northern Hemisphere spring snow cover since 1950. It is *very likely* that human influence has contributed to the observed surface melting of the Greenland Ice Sheet over the past two decades, but there is only *limited evidence*, with *medium agreement*, of human influence on the Antarctic Ice Sheet mass loss.

It is *virtually certain* that the global upper ocean (0–700 m) has warmed since the 1970s and *extremely likely* that human influence is the main driver. It is *virtually certain* that human-caused CO₂ emissions are the main driver of current global acidification of the surface open ocean. There is *high confidence* that oxygen levels have dropped in many upper ocean regions since the mid-20th century, and *medium confidence* that human influence contributed to this drop.



[xkcd \(a\)](#) showcases a very accurate prediction from [Exxon \(a\)](#), made back in 1982.

Kabul has fallen. One can laugh at Biden, and mention that he is "not a superforecaster". But forecasters and superforecasters alike also failed to see this one coming. I've written a postmortem from a forecasting perspective [here](#), available to Substack subscribers. Metaculus also has a post-mortem thread [here \(a\)](#).

[CDC recruits outsiders to lead a new center on disease forecasting \(a\)](#), with [Marc Lipsitch \(a\)](#) as Director of Science. The name might [ring a bell](#) for EA readers.

[5G Wireless Could Interfere with Weather Forecasts \(a\)](#). On the one hand, water absorbs electromagnetic radiation differently at different frequencies, and monitoring the 24 gigahertz frequency is apparently particularly informative. On the other hand, weather satellites use a 16 gigahertz band to communicate with stations on the ground. And proposed 5G bands might interfere with signals at either of those frequencies—reporting doesn't make clear which—and thus deteriorate weather forecasting performance.

...the biggest issue involves a spectrum called 24 gigahertz, which weather satellites use to monitor natural microwave signals produced by water vapor at various levels in the atmosphere. The device they use is a microwave radiometer.

But the signals made by water vapor and other natural weather signatures become fainter in a cacophonous surge of phone signals. "If you have a large network of cellphone towers transmitting many orders of magnitude more power near the ground,

some of that reflects upward and parts of the atmosphere will become very noisy," Mahoney said.

...the most "insidious" impact of rising noise levels on a weather spectrum would emerge if they caused errors or gaps in the weather data that is undetected. The erroneous data might be included in computer models that scientists use for, among other things, predicting future climate behavior.

Just where the FCC will go next with its Frontier Spectrum policy on 5G is unclear. According to the House Science Committee, it has already taken in almost \$2 billion from 29 winning bidders for space on the 24 gigahertz band.

The US military announces "Global Information Dominance" experiments, using machine learning to automate analyzing and collecting intelligence ([primary source \(a\)](#), [secondary source \(a\)](#)). Some highlights and thoughts [here \(a\)](#).

[Inflation Comes for Aluminum:](#)

Demand is set to surge on the back of climate-change investment, and mega-producer China—which accounts for more than half of global output—is cracking down on smelting to reduce pollution and meet green targets.

It's already jumped 26% this year to about \$2,500 a ton, one the best performers on the London Metal Exchange. Goldman Sachs Group Inc. is among those seeing more gains ahead, forecasting record prices above \$3,000 by late next year.

"It takes quite a mindset change—some viewed buying aluminum similar to buying groceries in the supermarket," said Philippe Mueller, head of aluminum trading at Trafigura. "It's not going to work like this anymore."

The metal isn't alone in facing short-term issues. The combination of soaring demand and spluttering supply after covid-19 disruption has upended many raw materials markets, all of which is feeding the global inflation scare that's taken hold in some corners this year.

On the topic of inflation, and to finish this newsletter with a forecast, see:

Will inflation be 0.4% or more from July to August?

74% Likely

● Last updated: 2021-09-01

This is a market on whether month-to-month inflation, as measured by the Bureau of Labor Statistics (CPI-U) will be 0.4% or more from July to August, when the BLS releases month-to-month inflation information for the preceding month of August on...



PolyMarket

Forecasts: 168

Liquidity: \$3.5k

Volume: \$12k

Source: [Will inflation be 0.4% or more from July to August?](#)

Note to the future: All links are added automatically to the Internet Archive, using this [tool](#) ([a](#)). "(a)" for archived links was inspired by [Milan Griffes \(a\)](#), [Andrew Zuckerman \(a\)](#), and [Alexey Guzey \(a\)](#).

Good ideas do not need lots of lies told about them in order to gain public acceptance

[Daniel Davies \(a\)](#)

Forecasting Newsletter: September 2021.

Highlights

- Facebook's Forecast [folds](#)
- [Hedgehog Markets](#) now operational
- Polygon and Augur announce a [\\$1M liquidity rewards program](#)

Index

- Prediction Markets & Forecasting Platforms
- Blog Posts
- Long Content
- In The News

Sign up [here](#) or browse past newsletters [here](#).

Prediction Markets & Forecasting Platforms

Hedgehog markets

[Hedgehog markets \(a\)](#) launches on the Solana mainnet, meaning that their prediction markets are now open to real-money bets. A brief explanation of how their no-loss competitions work and why they're interesting can be found on the [July edition of this newsletter \(a\)](#). Their first competition offers [\\$100k in prizes \(a\)](#).

Market	Date Range	Total Value Locked
[\\$100] NFL 2021 Weeks 3-5	SEP 22, 2021 - OCT 14, 2021	\$27,700
[\\$1000] NFL 2021 Weeks 3-5	SEP 22, 2021 - OCT 14, 2021	\$111,000
[\\$100] European Football 2021 Sept. - Oct.	SEP 22, 2021 - OCT 23, 2021	\$34,800
[\\$1000] European Football 2021 Sept. - Oct.	SEP 22, 2021 - OCT 23, 2021	\$119,000
[\\$100] Crypto Markets Fall 2021	SEP 22, 2021 - OCT 14, 2021	\$33,500
[\\$1000] Crypto Markets Fall 2021	SEP 22, 2021 - OCT 14, 2021	\$123,000

Hedgehog Markets' first [no-loss competitions](#)

Hedgehog's first markets are focused on sports and on the prices of various crypto assets. But their no-loss model is uniquely suitable for more long-term markets, so I'll be watching out for that.

Metaculus

[SimonM](#) (a) kindly curated the [top comments from Metaculus this past August](#). They are:

- [RyanBeck](#) (a) points out some shortcomings in Arnold Kling's fortified essay on [Two Theories of Inflation](#)
- [Charles](#) (a) looks at Macron's re-election chances in 2022
- [SimonM](#) (a) calculates a base rate for US government defaults linked to the debt ceiling using Laplace's law
- [ugandamaximum](#) (a) looks at Trump's electoral prospects in 2024
- [Comments on this question](#) (a) wrangle with how to measure AI damage
- [galen](#) (a) makes a well-informed comment on [What will be the Hue \(in angular degrees\) of Pantone's Color of the Year for 2022?](#) (a)

Metaculus is [advertising for various job openings](#) (a): Lead Developer, Senior NLP/Machine Learning Engineer, Effective Altruism Question Author, and “[analytical storyteller](#)” (a).

Polymarket

Polymarket writes about [Why You Should Get Your News From the Blockchain](#) (a) on one of the largest crypto newsletters. Meanwhile, Polymarket keeps experiencing outages because of troubles with The Graph—one of the services it uses to connect its webpage with its blockchain contracts.



Business

Will inflation be 0.4% or more from July to August?

Market ends on
September 14, 2021

Favorite

\$1.00

.75

.50

.25

.0

24h

7d

30d

All

Aug 18, 12AM Aug 27, 12PM Sep 06, 12AM Sep 15, 12PM Sep 25, 12AM Now

[Will inflation be 0.4% or more from July to August?](#), Polymarket

Last month, I lost a small amount of money following forecasts from [Star Spangled Gamblers \(a\)](#) and [Karlstack \(a\)](#), respectively on Britney Spears' conservatorship and August US inflation numbers. While I still find both sources very entertaining, I probably won't be following their lead on Polymarket in the future.

Hypermind

Hypermind launches another tournament: [Forecast Covid-19 Mortality Worldwide \(a\)](#), with \$15,000 in total rewards.

The header of the Le Point Politique website. It features the magazine's logo "Le Point" in large white letters on a red background, with "Politique" written below it. To the left is a menu icon and the word "MENU". Below the main title are several navigation links: Politique, International, Économie, Tech & Net, Culture, Débats, Sciences, Santé, Sports, Lifestyle, Afrique, Événements, and Abonnés. On the right side of the header is a small thumbnail image of the magazine cover.

Home • Politique

Hypermind – Zemmour devance Le Pen et Bertrand

Le marché prédictif présidentiel Hypermind- « Le Point » estime qu'Éric Zemmour a plus de chances de gagner l'élection que Marine Le Pen ou Xavier Bertrand.

Par Émile Servan-Schreiber (*)

Opinion piece on [Le Point](#) by Hypermind's Director

Hypermind was also featured on [Le Point](#) (a), a large French magazine, on the topic of their upcoming presidential elections. Per the article, Marine Le Pen's chances aren't as high as those of a more entertaining far-right candidate, [Eric Zemmour](#) (a). h/t MonsieurDimanche.

Odds and ends

[Polygon and Augur \(Turbo\) announce a \\$1M liquidity rewards program](#) (a).

The program incentivizes liquidity providers (LPs) by rewarding them through so-called liquidity mining, in which users of a decentralized finance (DeFi) product earn an additional token on top of the regularly expected yield just for putting assets into a liquidity pool. In return, liquidity made available helps bootstrap user adoption and ensure the smooth running of Augur Turbo. Users can earn rewards by providing liquidity to every side of the bet on the platform.

This is interesting because providing liquidity is one of the most fickle and unprofitable parts of running a prediction market. Historically, Polymarket—which is also on the Polygon chain—has been setting large amounts of their money on fire providing liquidity to their own markets.

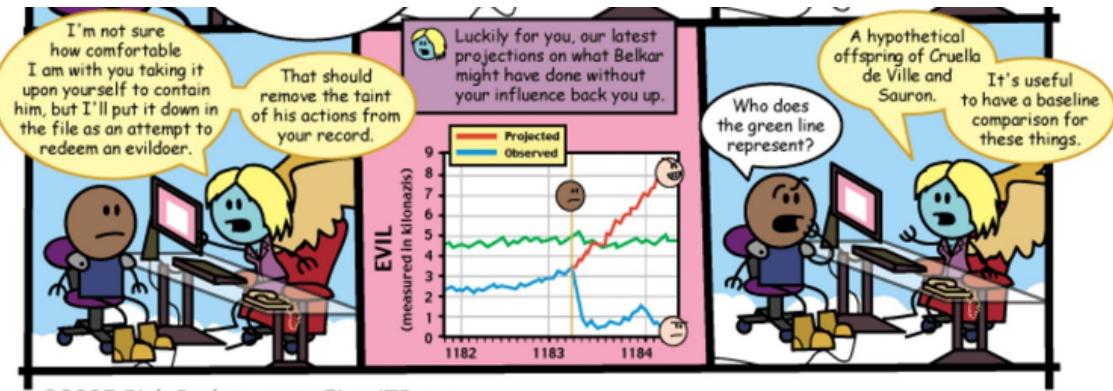
Augur Turbo is also setting their [trading fees to 0%](#) (a).

[Facebook's Forecast](#) (a), an experimental website and iPhone app from Facebook's New Product Experimentation division, [folds](#) (a). Second-hand sources speculate that this was because of the community's poor performance. Forecast was previously announced on LessWrong [here](#). h/t @dglid, Michał Dubrawski.

[PredictIt Arbitrage calculator](#) (a) fetches the PredictIt markets in which one can make money no matter what the outcome is—because the shares price doesn't sum to exactly 100%. If I'm reading this correctly, there aren't any salient opportunities right now.

Meanwhile in the corporate world, Amazon is searching for a [Network Forecasting and Planning](#) (a) Program Manager, with "Proficiency in MS Excel."

Blog Posts



©2007 Rich Burlew www.GiantITP.com

[The Order of the Stick](#), on quantifying the impact of a life

John Cochrane ([a](#)) writes [Climate Policy Should Pay More Attention to Climate Economics](#) ([a](#)):

But the best guesses of the economic impact of climate change are surprisingly small. The U.N.'s IPCC finds that a (large) temperature rise of 3.66°C by 2100 means a loss of 2.6 percent of global GDP. Even extreme assumptions about climate and lack of mitigation or adaptation strain to find a cost greater than 5 percent of GDP by the year 2100.

Now, 5 percent of GDP is a lot of money — \$1 trillion of our \$20 trillion GDP today. But 5 percent of GDP in 80 years is couch change in the annals of economics.

[...]

For a small donation, pictures of cuddly animals might do. For trillion-dollar costs and regulations, they do not. To justify such costs, we need some dollar value on specific environmental damage of climate change. Yes, the numbers are uncertain. But those numbers are the only sensible framework to discuss spending trillions of dollars on climate now.

I'll file this under "big if true". For what it's worth, the 2.6% of GDP is indeed mentioned in page 256 of [this IPCC report](#) ([a](#)):

Under the no-policy baseline scenario, temperature rises by 3.66°C by 2100, resulting in a global gross domestic product (GDP) loss of 2.6% (5–95% percentile range 0.5–8.2%), compared with 0.3% (0.1–0.5%) by 2100 under the 1.5°C scenario and 0.5% (0.1–1.0%) in the 2°C scenario

However, note that these are forecasts about events *80 years into the future*.

[Updates and Lessons from AI Forecasting](#) ([a](#)) gives [Jacob Steinhardt's](#) ([a](#)) outlook on having commissioned forecasts on the future of AI through Hypermind. See also [a forecaster's rundown of his predictions for Steinhardt's tournament](#) ([a](#)).

[Cultured meat predictions were overly optimistic](#) ([a](#)). "Of the 273 predictions collected, 84 have resolved - nine resolving correctly, and 75 resolving incorrectly. Additionally, another 40 predictions should resolve at the end of the year and look to be resolving incorrectly. Overall, the state of these predictions suggest very systematic overconfidence."

[When pooling forecasts, use the geometric mean of odds](#) ([a](#)). "There are many methods to pool forecasts. The most commonly used is the arithmetic mean of probabilities. However,

there are empirical and theoretical reasons to prefer the geometric mean of the odds instead." SimonM finds that [empirical Metaculus data](#) (a) confirms this.

[Measuring the information in an empirical prior](#) (a):

For example, clinical trials reporting hazard ratios for treatment effects of say $HR < 1/20$ or $HR > 20$ are incredibly rare and typically fraudulent or afflicted by severe protocol violations. And then an HR of 100 could represent a treatment for which practically all the treated and none of the untreated respond, and thus is far beyond anything that would be uncertain enough to justify an RCT – we do not do randomized trials comparing jumping with and without a parachute from 1000m up. Yet typical “weakly informative” priors assign considerable prior probability to hazard ratios far below 1/20 or far above 20.

Long Content

A flurry of papers [are out](#) (a) detailing results and conclusions from the [Makridakis 5 competition](#) (a), one of the largest and longest-running ML forecasting competitions in the world. They will be presented at the [M5 Conference](#) (a) this September. [Applicability of the M5 to Forecasting at Walmart](#) ((sci-hub link (a)) presents the scoring details, and compares Walmart's own data pipeline to the winning entries.

[Wikipedia: Predictions of the collapse of the Soviet Union](#) (a). "Whether any particular prediction was correct is still a matter of debate, since they give different reasons and different time frames for the Soviet collapse."

AI Impacts in [2015](#) (a), on a large dataset of predictions of when human-level AI will be achieved.

[Probabilistic Storytelling and Temporal Exigencies in Predictive Data Journalism](#) (a) ([sci-hub link](#) (a)) outlines some considerations about predictive storytelling:

- Better prediction-making capabilities have only been game-changing in a few journalistic industries (political predictions, weather forecasting, sports)
- The position of "data journalist" isn't very established; people with that position often call themselves by different names (journalists, researchers, data-analysts, interaction designers, etc.)
- Journalists are reluctant to use predictions
- This might be because
 - news are churned pretty fast, and focused on the very short term...
 - whereas predictions take time to gather and make sense of and tell stories around
 - or because text, stories and human minds tend to follow one thread...
 - whereas probabilistic futures are a garden of forking paths, and thus more difficult to represent

Personally, I feel that the paper doesn't push enough on the lack of incentives for numerical predictions, or on the disappointing inadequacy of most—but not all—journalists to work with data, predictions, or nuance more generally. Further, it doesn't propose actionable insights. For instance, this seems like an area which would be amenable to top-down improvement, e.g., by directly paying newspapers to invest in quality probabilistic journalism, or by establishing prizes to incentivize such journalism.

"Proper scoring rules, like log or Brier, incentivize an expert to report their true belief. But what if there are *multiple* experts? In that case they can collude to guarantee themselves a larger reward". See [this twitter thread](#) (a) or the accompanying [paper](#) (a).

[Hindenburg Research](#) (a)—a firm whose profit model is to investigate companies until it uncovers fraudulent activities, and then to short those while revealing their research—[alleged this June that DraftKings](#) (a), a major US betting operator, has and continues to deal in countries where betting is illegal.

In the News

[Nowcasting the Next Hour of Rain](#) (a). Google's DeepMind releases a more powerful model for forecasting weather in the very short term.

[Weather Forecasting in Afghanistan military operations](#) (a) is a nice example of superior forecasting prowess that had an example on the tactical level—e.g., being able to evacuate soldiers better, being able to better schedule attacks—but not at the strategic level—the US lost regardless.

FiveThirtyEight [challenges readers](#) (a) to beat their NFL forecasts.

[The Business of Forecasting Fashion](#) (a). In a Wall Street Journal podcast, a fashion forecasting expert talks about predicting what people will wear. What I found most interesting was trying to logically follow the impact of trends. For instance, as poorer millennials work more often from home, there is more relative demand for home wear.

Note to the future: All links are added automatically to the Internet Archive, using this [tool](#) (a). "(a)" for archived links was inspired by [Milan Griffes](#) (a), [Andrew Zuckerman](#) (a), and [Alexey Guzey](#) (a).

No airship will ever fly from New York to Paris. That seems to me to be impossible. What limits the flight is the motor. No known motor can run at the requisite speed for four days without stopping, and you can't be sure of finding the proper winds for soaring. The airship will always be a special messenger, never a load-carrier. But the history of civilization has usually shown that every new invention has brought in its train new needs it can satisfy, and so what the airship will eventually be used for is probably what we can least predict at the present.

Wilbur Wright, of the Wright Brothers, [1908](#) (a) h/t [Nintil](#) (a) through the [Best of Twitter](#) (a) newsletter.

Forecasting Newsletter: October 2021.

Highlights

- Polymarket is [being investigated](#) by the CFTC
- Metaculus solves long-lasting UX problem: it now allows more expressive distributions.
- David Friedman looks at the [track record of IPCC predictions](#)

Index

- Prediction Markets & Forecasting Platforms
- Job Board
- In the News
- Blog Posts
- Long Content

You can sign up for this newsletter [on substack](#), or browse past newsletters [here](#). If you have a content suggestion or want to reach out, you can leave a comment or find me on [Twitter](#).

Prediction Markets & Forecasting Platforms

Polymarket

Polymarket is [being investigated by the US Commodity Futures Trading Commission \(a\)](#). Back in 2008, several Nobel Prize winners and other academics [called on US agencies to clarify and establish regulations on prediction markets \(a\)](#). So at this point, I feel it's on the US Securities Exchange Commission and the Commodity Futures Trading Commission to have regulated this sooner, not on Polymarket not to have acted somewhat unilaterally.

Otherwise, Polymarket has continued to make more incremental usability improvements which make it more convenient for users to trade, add and withdraw funds from the site.

Metaculus

"nobody freak out but now you can asymmetrically adjust the tails of your logistic curves on continuous input questions on Metaculus" — @casens

- [SimonM \(a\)](#) kindly curated the [top comments from Metaculus this past August \(a\)](#). They are:
 - [SimonM \(a\)](#) calculates a market-implied probability of US default
 - [TeeJayKay \(a\)](#) gets deep into analysing Spotify's top songs
 - [Matthew_Barnett \(a\)](#) bets Robin Hanson that AGI will arrive before ems (and shares his logic)
 - [PeterWildeford \(a\)](#) goes long on Trump
 - [SimonM \(a\)](#) shares the historic General Social Survey data on sexlessness in young people
 - [NunoSempere \(a\)](#) lays out his logic for the rate of new donors appearing in the EA space
- Metaculus held a panel discussion on its [collaboration with the Virginia Health Department \(a\)](#):

Since March 2020, Metaculus has provided forecasting and modeling resources to public health professionals as they've made crucial decisions in tracking and combating COVID-19. These efforts include the ongoing Keep Virginia Safe and the recently concluded Virginia Lightning Round forecasting tournaments, which were developed in partnership with the Virginia Department of Health and the University of Virginia Biocomplexity Institute, and were designed to enhance COVID-19 modeling efforts while contributing to the ongoing public health policy conversation in Virginia.

I feel that there is a disconnect between the "crucial decisions" and the [very small prize pool of \\$1,000 given to forecasters](#) (a). The moment that Metaculus' questions influence public health policy in Virginia at all, their Department of Health's willingness to pay should go through the roof.

In collaboration with Rethink Priorities (and with Michael Aird in particular), Metaculus has started [short](#) (a) and [long-term](#) (a) nuclear risk tournaments.

Good Judgment

The best comments on Good Judgment Open for October, as curated by myself, were:

- [Kogo](#) (a) mentions that neither Europe nor China have the incentive to push for their Comprehensive Agreement on Investment in the current political standoff.
- [borisn](#) outlines why the Democrats will most likely [lose the House of Representatives](#) (a) but [retain control of the Senate](#) (a), based on historical frequencies.
- [dada](#) (a) wonders why the Good Judgment Open crowd assigns such a low probability to Iran and the US rejoining the JCPOA by the end of this year.
- [TerrySmith](#) (a) mentions that Good Judgment Open seems to have beaten PredictIt on the 2020 election forecasts, so looking at betting odds might not be that informative for the Good Judgment Open crowd when discussing the upcoming election in Virginia.
- [belikewater](#) (a) mentions a [declassified intelligence report on COVID](#) (a) which turned out to be inconclusive.

Good Judgment Open begins the [In The News](#) (a) and [Dubai Future Experts](#) (a) challenges.

Odds and ends

[Futuur](#) (a) is a prediction markets platform that recently came into my radar. They allow Americans to participate with a play-money currency, and the rest of the world to trade in dollars and in a variety of cryptocurrencies. They seem legitimate, and have been running a version of their site since 2017. But I would advise some caution: because all of their cryptocurrencies go into the same pool, Futuur could be on the hook if the price of any one cryptocurrency moves too much too fast.

Reddit [expands their prediction functionality](#) (a) (original sources: [1](#) (a), [2](#) (a)). I've also become aware of a prediction community of Reddit at [r/Predictor](#) (a). With 14.4k members on that community alone, Reddit might just have become one of the biggest prediction platforms around, almost without even trying. h/t [@marshallk](#) (a).

Augur continues to focus on sports betting on Polygon with Augur Turbo, and saw upwards of \$1M in trading volume, most likely because of the influx of [subsidized liquidity](#) (a).

CSET-Foretell continues to [make progress](#) (a) on their campaign around the [future of the relationship between the US Department of Defense and US tech companies](#) (a). They will have an event discussing their preliminary findings on the [10th of November](#) (a).

Hedgehog Markets concluded their first two competitions, and will launch new ones, still focused on crypto and sports. Kalshi has also made a few improvements on their desktop

webpage, and added some range markets, such as [this one \(a\)](#) on Chicago temperatures.

Forecasting Job Board.

The Perry World House (PWH) team at the University of Pennsylvania is looking for a [Program Manager for the Future of the Global Order \(a\)](#) because the current holder of the position is joining Founders Pledge. PWH has been "doing lots of work on implementing probabilistic forecasting methods in the U.S. government, and the person taking this job would likely continue work on those issues". One particularly high-quality piece of work by PWH previously mentioned in this newsletter was [Keeping Score: A New Approach to Geopolitical Forecasting \(a\)](#).

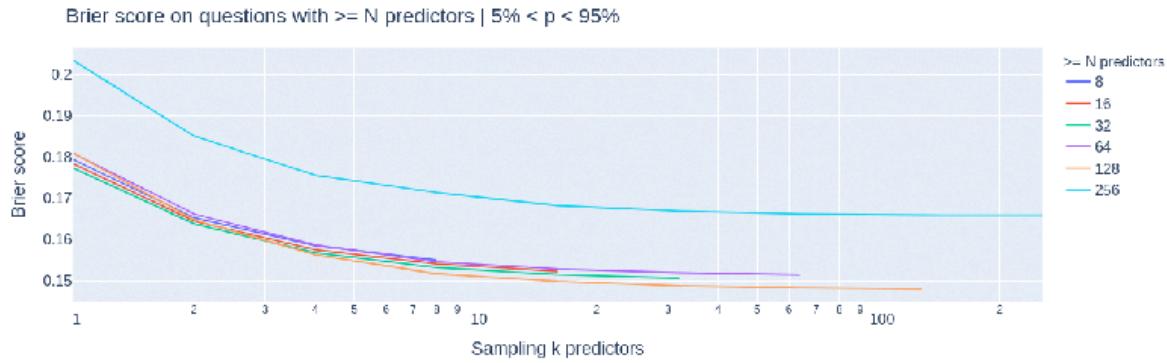
The Global Priorities Institute is looking for a [Research Assistant \(a\)](#) to aid its investigation into making forecasting a core research area. They are offering £17.48 (\$23.85, 20.62€) per hour.

Metaculus is searching for ["analytical storytellers" \(a\)](#) on a rolling basis, paying around \$0.3 per word ("essays are compensated at \$300 each and at \$25 per forecast question, with additional compensation awarded for especially high-quality essays attracting a significant readership")

North Dakota is looking for [economic forecasting consultants \(a\)](#). Although the offer seems to be aimed at individual consultants, I feel that it would also be interesting for forecasting platforms/prediction markets to apply.

Blog Posts

Charles Dillon of Rethink Priorities and SimonM look at [How does forecast quantity impact forecast quality on Metaculus? \(a\)](#). More forecasters increase forecast quality, but the effect is small beyond 10 or so forecasters.



Forecasting performance as a function of the number of predictors, by [SimonM](#) using Metaculus data.

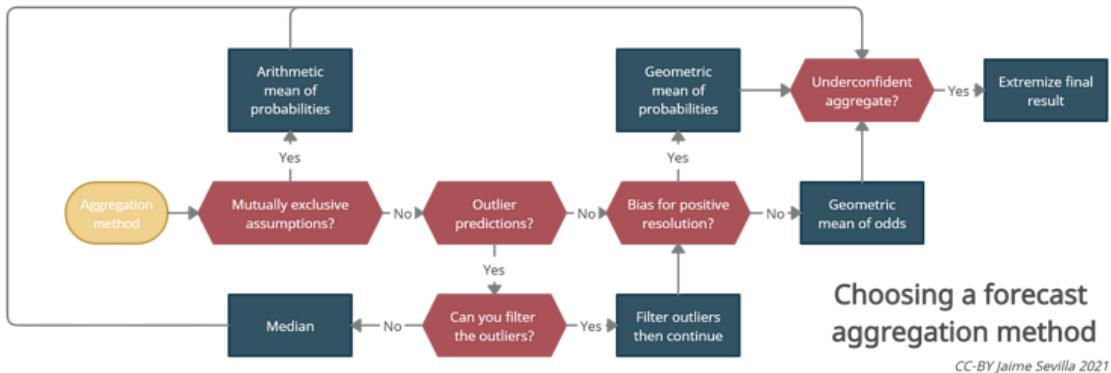
Forecasting performance as a function of the number of predictors, by [SimonM](#) using Metaculus data.

One possible driver of this effect could be Metaculus allowing up to 10 forecasters to meaningfully coordinate in the public comments section, but not much beyond that.

David Friedman looks at whether the [past IPCC temperature projections/predictions have been accurate? \(a\)](#)

The predictions look better now than they did in 2014, high three times out of four, low once, and only once has actual warming been below the predicted range. They are still running a little high but the results look consistent with random error. That makes it at least possible that the IPCC researchers are now modeling the climate system well enough to produce reasonable estimates of its future behavior.

Jaime Sevilla writes about his [current best guess on how to aggregate forecasts \(a\)](#):



In [Learning from our \(the USA's\) defeat \(a\)](#) Tanner Greer of The Scholar's Stage looks at the leadership team of the second Bush's administration. It seems very much worth reading in terms of improving one's models of the world.

A [Salesforce blogpost \(a\)](#) advertises the wonders of cloud-based enterprise resource planning solutions (such as Salesforce itself.) Nonetheless, they still know what they are talking about.

In the News

Quartz covers the [shutdown of Facebook's Forecast \(a\)](#) in more depth.

[FiveThirtyEight \(a\)](#) writes a data-driven analysis of Biden's approval ratings.

[US car sales are expected to plummet \(a\)](#) due to chip shortage. I keep seeing this term "chip shortage", but it seems to me that this is more of a "supply chain mismanagement" issue because chips alone can't really make up that high a relative proportion of a vehicle's prize. Not also that Tesla doesn't seem to be affected by this shortage.

[Warmer-than-normal temperatures could help save Americans on home heating costs, which could be elevated this year due to high energy prices \(a\)](#), reports CNN. I find this curious because I'd expect CNN to avoid mentioning anything that could suggest that climate change is not unalloyedly negative. Still, the article doesn't mention the impact of Biden on energy prices, nor climate change directly.

Sephora, a beauty products brand, [integrates AI more into its forecasting and replenishment software \(a\)](#). To be clear, this is just business as normal, but it still feels like the kind of thing which is more likely in a [world with short AI timelines \(a\)](#).

Long Content

[Issues with Futarchy \(a\)](#) compiles possible failure modes with a governance model proposed by Robin Hanson where decisions would be made based on prediction markets.

Note: Due to [EA Global](#) (a), I now have a backlog of forecasting [effort posts](#) (a) posted during October. They will be incorporated into the next edition of this newsletter.

Note to the future: All links are added automatically to the Internet Archive, using this [tool](#) (a). "(a)" for archived links was inspired by [Milan Griffes](#) (a), [Andrew Zuckerman](#) (a), and [Alexey Guzey](#) (a).

Are you Alex Lawsen? *The [Alex Lawsen](#)*?

— Anonymous, EA Global 2021.

Forecasting Newsletter: November 2021

Highlights

- Polymarket sees [record-high swings](#)
- Replication Markets [pays out \\$142k in forecaster rewards](#)
- The Economist features a [full page](#) with Good Judgment Open's forecasts

Index

- Prediction Markets & Forecasting Platforms
- In The News
- Blog Posts
- Long Content

Sign up [here](#) (a) or browse past newsletters [here](#) (a).

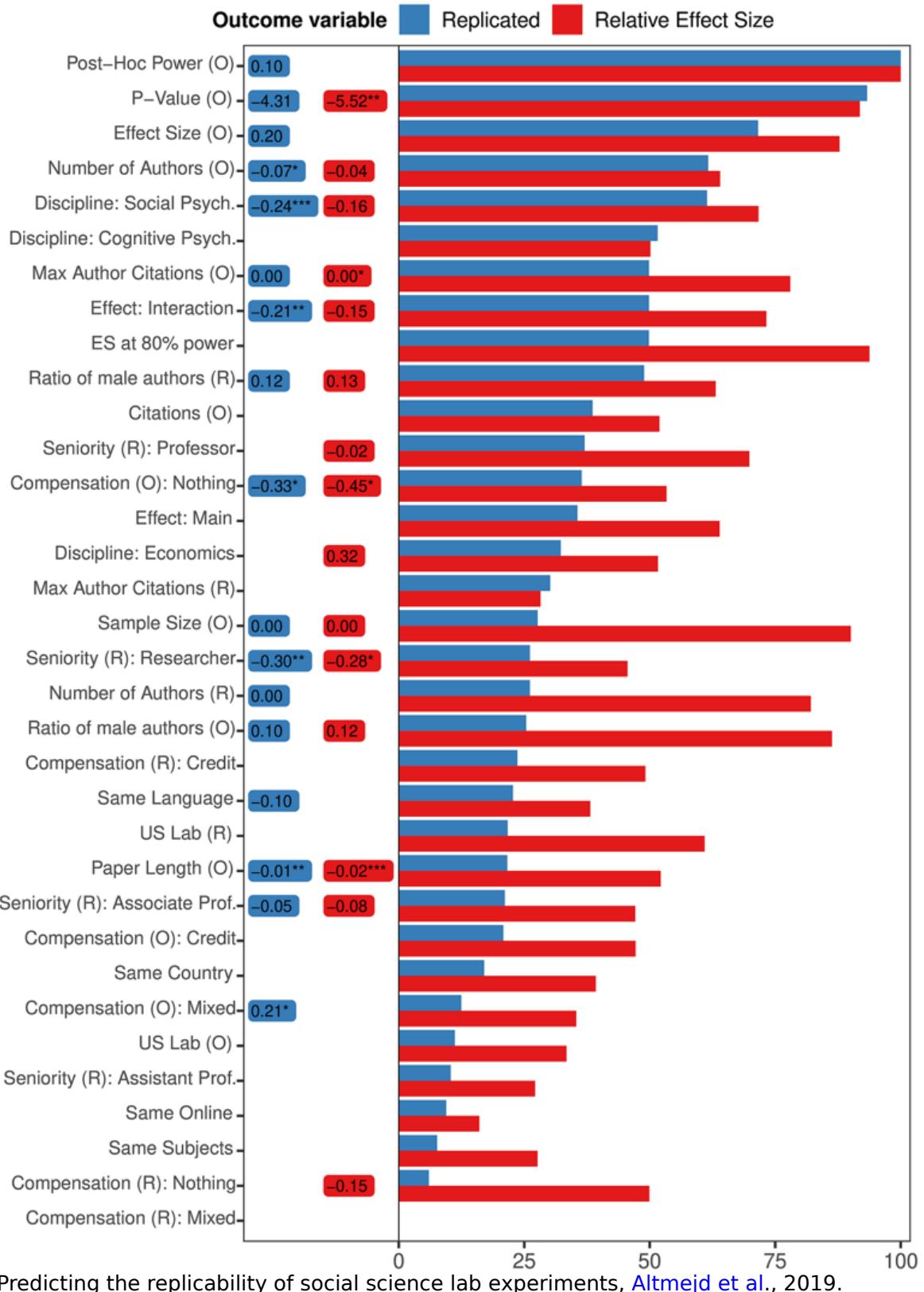
Prediction Markets & Forecasting Platforms

Replication Markets

Replication Markets (a) was a project to research how well prediction markets could predict whether papers would replicate. They are [paying out](#) (a) \$142k in cash rewards for the prediction markets part of the experiment. This corresponds to 121 resolved questions, which includes 12 meta-questions and 30 about covid papers.

The leaderboard for users is [here](#) (a). I won a mere \$809, and I don't remember participating all that much. In particular, I was excited at the beginning but lost interest because a user—or a bot—named "unipedal" seemed like it was taking all the good opportunities.

Now, a long writeup by "unipedal" himself can be read at [How I Made \\$10k Predicting Which Studies Will Replicate](#) (a). The author started out with a simple quantitative model based on [Altmejd et al. \(2019\)](#) (a)



Predicting the replicability of social science lab experiments, [Altmejd et al., 2019](#).

But in later rounds, he dropped the quantitative model, and started "playing the market". That is, he found out that trying to predict how the market will move is more profitable than

giving one's own best guess. Unipedal then later automated his trades when the market API was opened to users.

In contrast, I participated in a few rounds and put in 10x less effort while earning much more than 1/10th of the rewards. As unipedal points out, this is backwards:

...I think one of the most important aspects of "ideal" prediction markets is that informed traders can compound their winnings, while uninformed traders go broke. The market mechanism works well because the feedback loop weeds out those who are consistently wrong. This element was completely missing in the RM [Replication Markets] project.

The same author previously wrote: [What's Wrong with Social Science and How to Fix It: Reflections After Reading 2578 Papers](#) (a), which is also based on his experiences with the Replication Markets competition.

Besides Replication Markets, DARPA has also founded another group to predict replications through their [SCORE](#) (a) program. Based on preliminary results, this [second group](#) (a) seems like they beat Replication Markets by using a more [Delphi-like](#) (a) methodology to elicit predictions.

Metaculus

It has been an active month for Metaculus.

For starters, they [rehauled](#) (a) their scoring system for tournaments. Then, Metaculus' [Journal](#) (a) started to give fruit: The article on forecasts of [Human-Level Language Models](#) (a) (also on LessWrong [here](#) (a)) was of fairly high quality.

Metaculus also started to keep track of the accuracy of a small number of [Public Figures](#) (a). Because Metaculus has so many questions, every time one of these figures makes a public prediction, it is likely enough that Metaculus also has a prediction on the same issue. Over time, this will allow Metaculus to see who is generally more accurate. This is a more adversarial version of Tetlock's original [Alpha Pundit](#) (a) idea: instead of having experts willingly participate, Metaculus is just passively keeping track of how bad they are. Kudos!

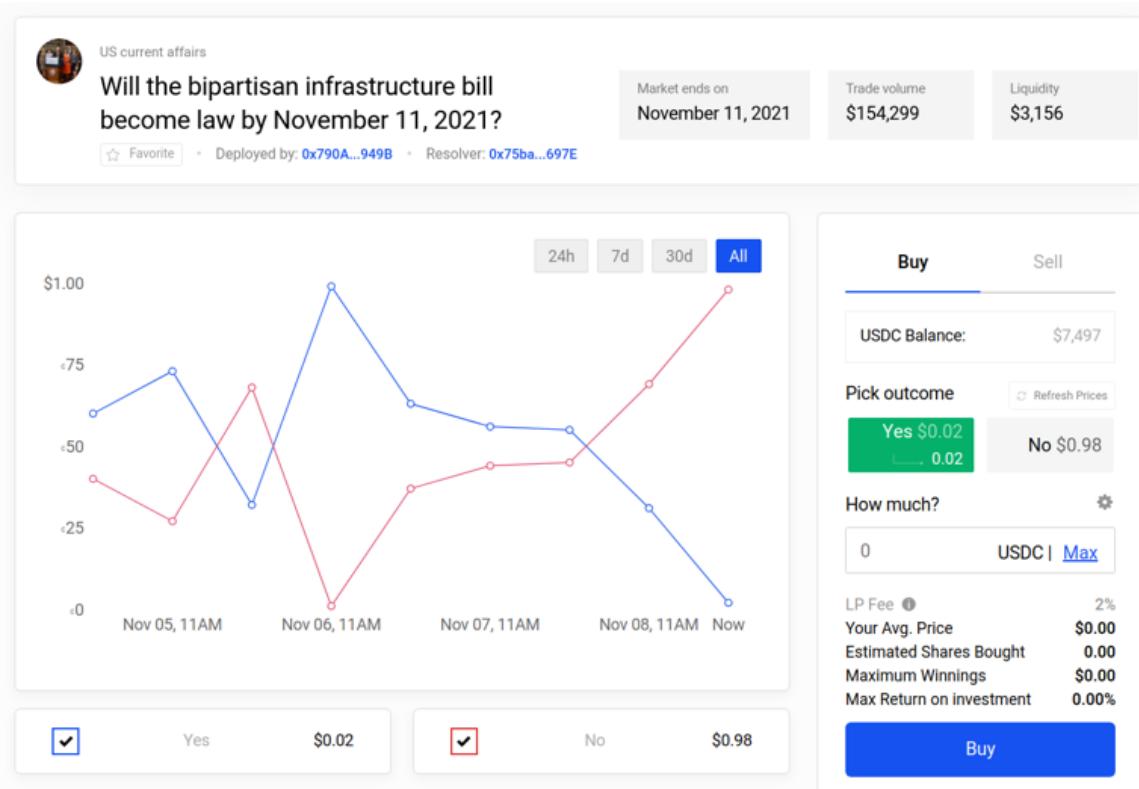
Two comments worth highlighting from [SimonM](#)'s (a) list of [top comments from Metaculus this past November](#) (a) are:

- [juancambeiro](#) (a) re-opens a previously closed question on whether or not a member of the IC community thinks COVID was a lab leak.
- [ege_erdil](#) (a) thinks we should be extremely uncertain about crime. "Overall I think everyone in this thread is way too confident that they know what's going on with crime rates at some frequency scale. My opinion is that let alone understanding the long-term mechanisms which drive changes in crime rates, we don't even have a very good understanding of crime rates from the past. If Louis XIV's reign in France cut murder rates in half, we would never know it from the evidence available to us today."

In addition, Metaculus and its community worked at full speed to put up questions and produce forecasts on the [Omicron](#) (a) [variant](#) (a). Metaculus also added more questions to the [Keep Virginia Safe Tournament](#) (a), and increased the price pool somewhat to \$2,500.

Polymarket

Polymarket saw some extremely large swings, where [1:250](#) (a) and [1:700](#) (a) underdogs ended up winning. h/t [@Domahhhh](#) (a)



Zvi positively covers some Polymarket markets on Covid [here](#) (a).

Polymarket also added support for [Metamask](#) (a), one of the most popular crypto-wallets, making Polymarket ever more mainstream. They also had a bit of a brouhaha on a market on the number of [exoplanets](#) (a) discovered, where the resolution source pointed to two different numbers.

Odds and ends

Augur—a set of pioneering prediction market contracts on Ethereum and the community around it—is creating a [decentralized autonomous organization \(a\)](#), [AugurDAO \(a\)](#). I get the impression that the original developers have gotten tired of supporting Augur, whose current focus on sports markets merely makes it a very slow sportsbook.

But the move is also consistent with Augur's initial ethos of being decentralized. For example, the [Forecast Foundation \(a\)](#) which supports Augur's development, seems to live under [Estonian jurisdiction \(a\)](#), whereas a DAO arguably lives under no jurisdiction.

The Foresight Institute is hosting a "[Vision Weekend \(a\)](#)" in the US and France. Although I remembered the Foresight Institute as [something that Eric Drexler founded before he went on to do other things \(a\)](#), I did find some familiar names in the [list of presenters \(a\)](#), and browsing the details the event is probably going to be of higher quality than I would have thought.

The Anticipation Hub is hosting a "[Global Dialogue Platform on Anticipatory Humanitarian Action \(a\)](#)", hosted online from the 7th to the 9th of December. Although it seems more focused on global health and development topics, it might be of interest to NGOs around the forecasting space more generally.

[The Global Priorities Institute \(a\)](#) is dipping its toes into forecasting. As one might expect, so far there is a lot of academically-flavored discussions, but very little actual forecasting.

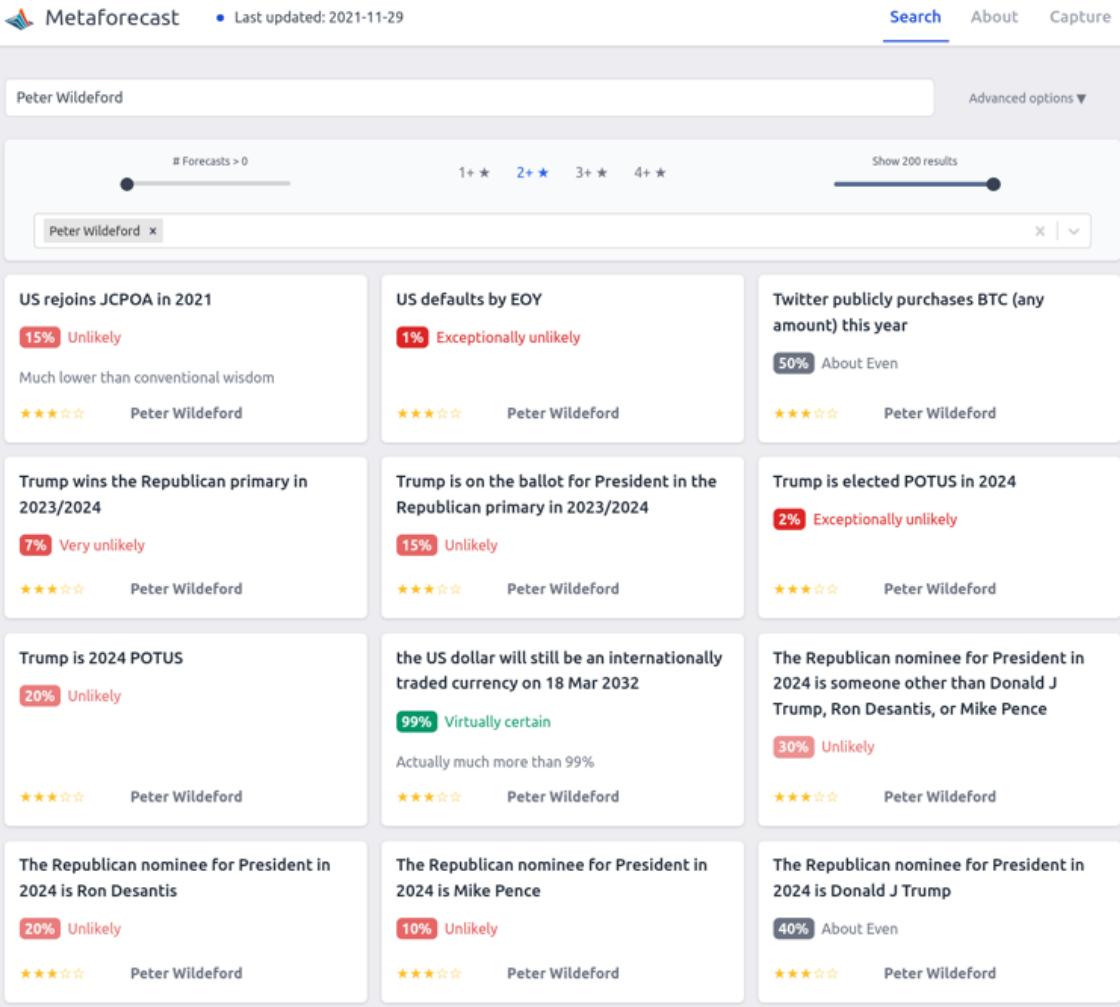
Hedgehog Markets had an [NFT-minting event \(a\)](#), where users could buy NFTs which they will be able to use to participate in competitions closed-off to non-NFT holders. I don't see the appeal, but others did and spent around \$500k on these tokens (5000 NFTs at [0.5 SOL \(a\)](#) per token).

A former US [Commodity Futures Trading Commission](#) commissioner [joined Kalshi's board \(a\)](#).

Hypermind started a [new contest \(a\)](#) on the "future of Africa", with \$6000 in prize money.

There is a [fairly neat calibration app \(a\)](#) based on the exercises from The Scout Mindset.

I've added [Peter Wildeford's publicly available predictions](#) to [Metaforecast](#):



On the negative side, Metaforecast is experiencing some difficulties with updating with new forecasts, which I hope to get fixed in the coming week.

In the News

The Economist featured a [full page with forecasts from Good Judgment Open \(a\)](#) on their "The World in 2022" edition.

Reuters reports that [Climate change extremes spur U.N. plan to fund weather forecasting \(a\)](#). My impression is that climate change fears are being used to fund much-needed bog-standard weather forecasting. I have mixed feelings about this.

An Excel competitor with some forecasting functionality, [Pigment \(a\)](#), raises [\\$73M \(a\)](#).

Blog Posts

Joe Carlsmith writes down his thoughts on [Solomonoff induction \(a\)](#) (see a decent introduction of the concept [here \(a\)](#)). Although I was already familiar with the concept, I still feel like I learnt a bunch:

- the [speed prior](#) (a) is a nice hack to get around the fact that some programs can run forever, and you can't say which ones they are per the Halting problem.
- There is some unavoidable sense in which one has to assign smaller probabilities to longer programs.
- Solomonoff Induction requires that uncomputable processes like Solomonoff Induction be impossible. If the universe could include uncomputable processes, it could include a copy of your Solomonoff Induction process. In that case, the universe could function as a Solomonoff "anti-Inductor". That is, the universe could perfectly simulate what your Solomonoff Inductor will predict next and then feed you the opposite.

Tanner Greer of The Scholar's Stage has a new piece on [Sino-American Competition and the Search For Historical Analogies](#) (a). His main point is that the tensions around Taiwan break the analogy between the current relationship between the US and China and the relationship between the US and the USSR during the Cold War.

Jaime Sevilla posts [A Bayesian Aggregation Paradox](#) (a): There is no objective way of summarizing a Bayesian update over an event with three outcomes A:B:C as an update over two outcomes A: \neg A. From the comments:

Imagine you have a coin that's either fair, all-heads, or all-tails. If your prior is "fair or all-heads with probability 1/2 each", then seeing heads is evidence against "fair". But if your prior is "fair or all-tails with probability 1/2 each", then seeing heads is evidence for "fair". Even though "fair" started as 1/2 in both cases. So the moral of the story is that there's no such thing as evidence for or against a hypothesis, only evidence that favors one hypothesis over another.

Long Content

[Are "superforecasters" a real phenomenon?](#) (a). David Manheim, a superforecaster, answers:

So in short, I'm unconvinced that superforecasters are a "real" thing, except in the sense that most people don't try, and people who do will do better, and improve over time. Given that, however, we absolutely should rely on superforecasters to make better predictions than the rest of people - as long as they continue doing the things that make them good forecasters.

Note to the future: All links are added automatically to the Internet Archive, using this [tool](#) (a). "(a)" for archived links was inspired by [Milan Griffes](#) (a), [Andrew Zuckerman](#) (a), and [Alexey Guzey](#) (a).

It is curious to reflect that out of all the "experts" of all the schools, there was not a single one who was able to foresee so likely an event as the Russo-German Pact of 1939. And when news of the Pact broke, the most wildly divergent explanations were of it were given, and predictions were made which were falsified almost immediately, being based in nearly every case not on a study of probabilities but on a desire to make the U.S.S.R. seem good or bad, strong or weak. Political or military commentators, like astrologers, can survive almost any mistake, because their more devoted followers do not look to them for an appraisal of the facts but for the stimulation of nationalistic loyalties

— George Orwell, [Notes on Nationalism](#) (a), 1945, h/t Scott Alexander.

Forecasting Newsletter: December 2021

Highlights

- Polymarket's future is uncertain after it [settled with the CFTC for \\$1.4M](#)
- Astral Codex Ten gives out \$40k to [forecasting projects](#)
- Many people, including [Mathew Yglesias](#), write predictions for 2022.
- Eli Lifland writes *the* reference piece on [bottlenecks to impactful forecasting](#)
- Google reveals the existence of a gigantic new [internal prediction market](#)
- A new forecasting platform appears, [Manifold Markets](#)

Index

- Prediction Markets & Forecasting Platforms
- Long Content
- Blog Posts
- In The News

You can sign up for this newsletter on [substack](#), or browse past newsletters [here](#). If you have a content suggestion or want to reach out, you can leave a comment or find me on [Twitter](#). A big hat tip goes to Nathan Young and Clay Graubard for comments and suggestions on this edition.

Prediction Markets & Forecasting Platforms

Polymarket

The US Commodity Futures Trading Commission (CFTC) has [fined Polymarket \\$1.4M \(a\)](#). For reference, Polymarket's seed funding amounted to [\\$4M \(a\)](#) and were in talks for another round at a [nearly \\$1B valuation \(a\)](#) prior to the investigation.

The order requires that Polymarket pay a \$1.4 million civil monetary penalty, facilitate the resolution (i.e. wind down) of all markets displayed on Polymarket.com that do not comply with the Commodity Exchange Act (CEA) and applicable CFTC regulations, and cease and desist from violating the CEA and CFTC regulations, as charged.

With this, Polymarket's future seems now very uncertain. The Polymarket team has been mostly silent, though it recently released an [update \(a\)](#) as a Google doc, which promised that they were just "getting started".

To quantify this uncertainty, I asked a more experienced prediction market trader—who wishes to remain anonymous—for his probability estimate that Polymarket would be "pretty much dead". This was operationalized as there being no new markets with more than \$100k in volume by November 2022. **His guess ranged from 30% to 50% that it'd in fact be dead.**

This is terrible news. Polymarket has recently been one of the very few real-money markets where one could find highly liquid markets on things of actual importance, like the covid pandemic. I don't tire of mentioning this [;2008! piece \(a\)](#) by several Nobel prize winners and other notable figures urging the CFTC to essentially allow these markets to exist.

[As an aside, I'll elaborate a bit on the method of elicitation, because I thought it was bloody ingenious, and it might be more standard in the future. I offered him \$100 to suggest a bet

on Polymarket being dead where I can take either side. He selects the odds, and is also allowed to have a spread, i.e., a difference between the bets against and the bets in favor. Then, I can choose to take the bet on either side, or to leave at that. In effect, I was paying him to potentially act as a bookie, where this costs him some effort and produces some probabilities I'm interested in as a side effect.

As a result of this process, he offered me bets ranging from 3:7 to 1:1, corresponding to between a 30% and a 50% probability of Polymarket being dead. So his revealed confidence was that the probability of it happening is in between.

This process was messier and involved some negotiation. But in the future, conditions could be standardized, e.g., instead of “a few thousands”, either side’s maximum bet could be predetermined to be \$5k, as we finally agreed. One would also have to think about the policy around making this kind of bet public. In particular, there are some issues around adversarial selection, or insider trading. If I or others had some private information about whether Polymarket was going to survive, I could extract money from the other party.]

CSET-Foretell

[CSET-Foretell](#) (a) is a forecasting platform that aimed to produce policy-relevant predictions and insights to influence US policy.

Foretell is now moving from being hosted by the [Center For Security and Emerging Technologies](#) (a) (CSET) at Georgetown University, to being hosted by the ["Applied Research Laboratory for Intelligence and Security"](#) (a) (ARLIS) at the University of Maryland.

The University of Maryland is generally less prestigious than the University of Georgetown, but Maryland ranks in the top 5 for [Homeland Security Graduate Schools](#) (a). ARLIS is also one of a very few Department of Defense [University Affiliated Research Centers](#) (UARCs) (a) and the only one in the [DC-Metro Area](#) (a).

I feel bitter about this, because I had high hopes for the platform, and because I expect ARLIS to be worse than CSET according to my values. On the one hand, CSET has received \$100M from [OpenPhilanthropy](#) (a) within a few years, whereas organizations similar to ARLIS historically receive [one million to \\$50M a year](#) (a) (page 55), and I'd expect ARLIS to receive an amount on the lower end of that range.

I'd also prefer funding which, broadly speaking, cares about people generally—like that of OpenPhil—over funding from the Department of Defense—which I'd expect would be more focused on the interests of the US alone.

On the positive side, ARLIS seems more deeply enmeshed into the US government's bureaucracy, and may [have the ear](#) (a) of [Kathleen Hicks](#) (a), a high-ranking US government official.

CSET's Michael Page also published [Wisdom of the Crowd as Arbiter of Expert Disagreement](#) (a), which outlines a methodology for using forecasts to resolve policy debates.

As in the first season, Samotsvety Forecasting, a team made up of Eli Lifland, Misha Yagudin, and myself, completely demolished the competition. We were around twice as good as the next-best team in terms of the relative Brier score.

Rank	Team	Scored Questions	Overall Score
1.	Samotsvety Forecasting	24	-1.105
2.	Ariana's Team	24	-0.586
3.	Olavo's Team	13	-0.556
4.	Aisha's Team	14	-0.49
5.	Pro Forecasters	16	-0.477
6.	Wake Forest University	6	-0.384
7.	Saaima's Team	13	-0.178
8.	Shaun's Team	24	-0.166
9.	Mariana's Team	4	0.046
10.	Molly's Team	12	0.076

Amusingly, all three of us are in the top 3 of all time (out of 1035 contenders.)

Rank	Username	Scored Questions	Overall Score
1.	 elifland	49	-1.53
2.	 Loki	35	-1.443
3.	 yagudin	33	-1.217
4.	 mollygh	17	-1.15
5.	 JS	15	-1.149
6.	 TrishBytes	17	-1.107
7.	 page	20	-1.025
8.	 JJMLP	49	-1.001
9.	 shaun-ee	30	-0.84
10.	 Zain_15	5	-0.784

Metaculus

[SimonM](#) highlights some comment threads [from Metaculus this past December](#) ([a](#)). They are:

- Discussion on [Roe vs Wade](#) ([a](#)), [part 2](#) ([a](#)), [part 3](#) ([a](#))
- Discussion on whether or not [Russia will invade Ukraine in 2022](#) ([a](#)), [part 2](#) ([a](#)), [part 3](#) ([a](#))
- Discussion on [how dangerous Omicron is](#) ([a](#)), [part 2](#) ([a](#)), [part 3](#) ([a](#))

Otherwise, the Economist partnered with Metaculus for a [Global trends in 2022](#) ([a](#)) tournament. Tom Chivers wrote a piece on solar power [for Metaculus](#) ([a](#)).

Odds and ends

Hedgehog Markets announced their [liquidity provider program \(a\)](#). The idea is to use the money which prediction market participants park to generate some return, which can then be given out as a reward to the best predictors. And Hedgehog Market is looking for partners, e.g., other Solana protocols, to generate that return for them.

Although the idea is interesting and innovative, [Hedgehog Markets \(a\)](#) continues to focus on sports, crypto, e-sports and NFT markets. And I view these topics as not being all that valuable to predict.

Still, Hedgehog Markets has the advantage that it allows participants to bet without losing their money. If the participant finally withdraws their money after a time, the scheme could be viewed as somewhat similar to a [tontine \(a\)](#), or to a [susu \(a\)](#), i.e., as a very simple savings device.

[Astral Codex Ten \(a\)](#) awarded \$1.55 million in grants, of which \$40k (2.5%) went to forecasting related projects. These were:

- [James Grugett](#), Stephen Grugett and [Austin Chen](#), \$20,000, for a new prediction market—Manifold Markets. If every existing prediction market is Lawful Good, this team proposes the Chaotic Evil version: anyone can submit a question, questions can be arbitrarily subjective, and the resolution is decided by the submitter, no appeal is allowed
- [Nikos Bosse](#), \$5,000, to seed a wiki about forecasting
- [Nathan Young](#), \$5,000, to fund his continued work writing Metaculus questions and trying to build bridges between the forecasting and effective altruist communities
- Nuño Sempere (myself), \$10,000, to fund his continued work on [metaforecast.org](#) and the [@metaforecast](#) bot.

[Manifold Markets \(a\)](#) was previously called Mantic Markets because of a section on Astral Codex Ten named "Mantic Mondays", but recently changed its name to allow that section to remain impartial. I find the platform rather neat.

In particular, most other prediction markets/forecasting platforms—like Hypermind, CultivateLabs, Betfair, Metaculus, etc.—were mostly programmed in or before 2015, with the web technologies of the time. Unlike them, Manifold Markets looks and feels more modern, which is something that I appreciate a lot as a heavy user of various forecasting platforms.

Moreover, the team has a couple of ex-Googlers, so beating everyone else in the technology front seems like a plausible pathway to dominance. I encourage people to [give it a try \(a\)](#). Some of the markets are entertaining, and for now, it's just play money.



🔥 Markets

Will Bitcoin be worth less than \$40,000 at some point before Jan 10, 2022 at 12 am GMT?

@PeterHrošo • M\$ 4,096 pool

48%
chance

Will Mantic Markets have 5 paying users by Feb 01, 2022?

@AustinChen • M\$ 1,275 pool

88%
chance

If Scott Alexander creates a market, will 100 people have traded on Mantic by Feb 1, 2022?

@AustinChen • M\$ 1,026 pool

95%
chance

Will February 2022 start right after January 2022?

@ChristianSterr • M\$ 257 pool

90%
chance

Search markets

By creator ▾

For those curious, an explanation of Manifold Markets' tricky dynamic parimutuel betting system can be found [here \(a\)](#).

In the interest of transparency, and because I think it's interesting in its own right, my application can be found [here \(a\)](#).

I applied to a large extent because [Nathan Young](#) specifically was cheerleading the embryonic capabilities of the current [@metaforecast](#) bot, and proposing new things that could be built on top of it. I realized that server costs could stack up fairly quickly, particularly in the best case scenario where a lot of people use it, as in the case of @threaderapp. And if ACX wanted to bankroll a weekend project of mine, why not.

With the benefits of hindsight, I should also have applied for more money and for more ambitious projects, and for the [Quantified Uncertainty Research Institute \(a\)](#), the org for which I work, rather than for hobby projects. But I forgive myself, because initially this was going to be a \$200k grant round, before Vitalik Buterin and others bumped it up to \$1.55M.

Lastly, it's kind of interesting how \$40k feels like a significant quantity of all the funding there is for small experiments in the forecasting space. This is probably suboptimal.

Nathan Young and I are organizing a chill online meetup at 7:00 PM [UTC](#) on the [9th of February in the LessWrong online Walled Garden](#). The LessWrong Walled Garden is great because you can leave and join conversations as you wish, allowing better flowing conversations. The event will officially finish 2 hours after it starts, but anyone is welcome to stay later

Long Content

Eli Lifland publishes what is now [the reference piece on bottlenecks to more impactful forecasting \(a\)](#). It crystallizes his knowledge from a few years of his forecasting on

Metaculus, CSET-Foretell and Good Judgment Open.

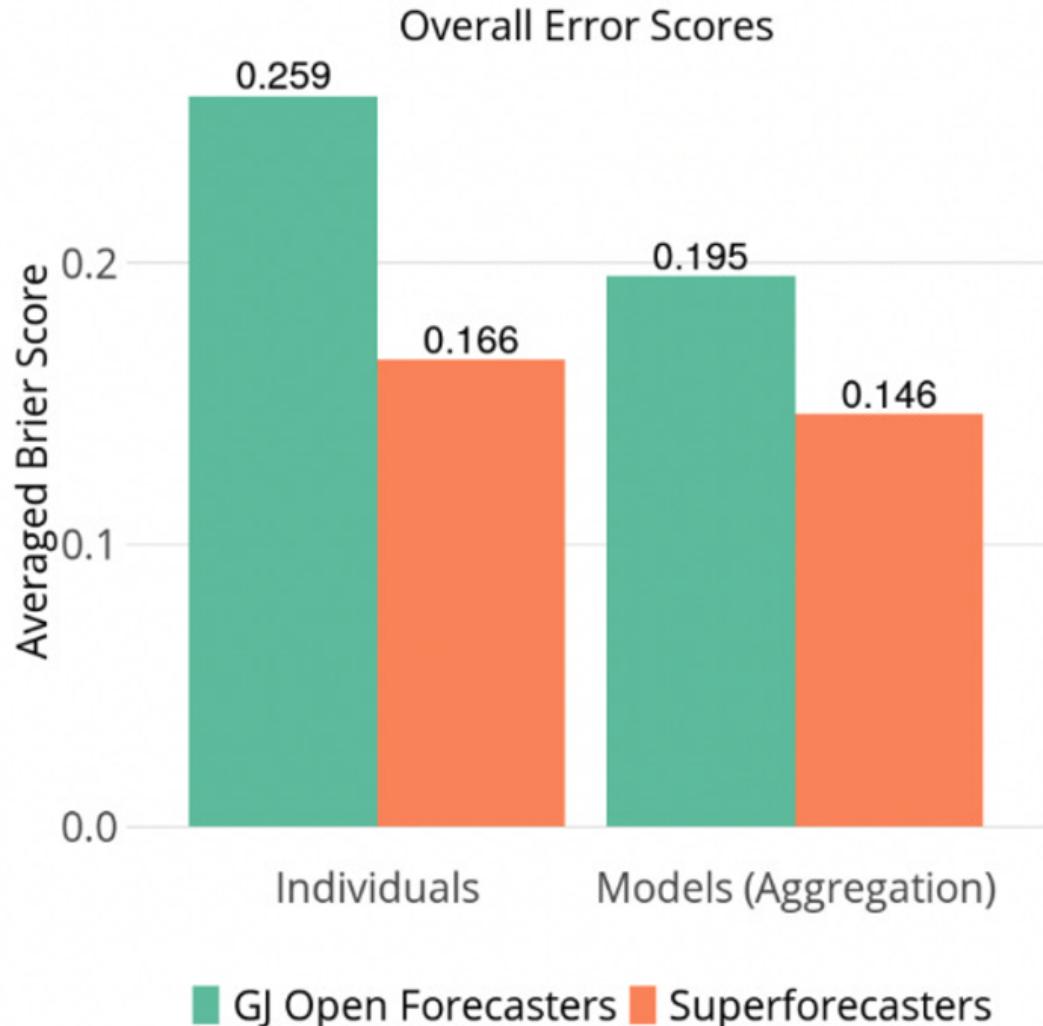
The data from the original Good Judgment Project is [available publicly](#) ([a](#)), and has been for some time.

Together with my co-authors Misha Yagudin and Eli Lifland, I posted a fairly thorough investigation into [Prediction Markets in The Corporate Setting](#) ([a](#)). The academic consensus seems to overstate their benefits and promisingness. Lack of good tech, the difficulty of writing good and informative questions, and social disruptiveness are likely to be among the reasons contributing to their failure. In the end, our report recommended not having company-internal prediction markets.

Dan Schwarz, who leads the below-mentioned prediction market at Google, answered on Twitter [here](#). Ozzie Gooen, of my own Quantified Uncertainty Research Institute, left some criticisms [here](#) ([a](#)), emphasizing that small experiments may nonetheless be worth it, and that experimenting with prediction setups could have large externalities if they work.

A [white paper by the Good Judgment project](#) ([a](#)) compares the performance of superforecasters vs Good Judgment Open forecasters. I recommend mostly skipping the text because of the information is contained in the charts.

The paper also doesn't have the visceral impact of the comparison in the Superforecasting book, where the original Good Judgment project beat intelligence analysts with classified information. This time, the paper compares paid superforecasters against unpaid hobbyists. I guess I'd have liked to see a comparison between different platforms, e.g., a Good Judgment vs Metaculus or vs PredictIt head-to-head fight.



Jaime Sevilla looks at [aggregating forecasts in a principled way](#) (a), building on his [previous work](#) (a). This time, he explains [a result by Neyman et al.](#) (a), and tests it on past Metaculus data. He finds that it beats Metaculus' own prediction, as well as all other aggregation methods commonly considered.

Blog Posts

The Machine Intelligence Research Institute has published a few [conversations on future AI capabilities](#) (a). Of these, readers of this newsletter might be particularly interested in the [Conversation on technology forecasting and gradualism](#) (a).

[Epidemic tracking and forecasting: Lessons learned from a tumultuous year](#) (a) summarizes a collection of papers on the PNAS (Proceedings of the National Academy of Science). The main lessons are:

1. Data was often unreliable
2. It is important to understand what process generates the data.
3. Mandated reporting was burdensome and inflexible

4. Human behavior was hard to model.

See also Valentine's [What are sane reasons that Covid data is treated as reliable? \(a\)](#).

[A non-magical explanation of Jeffrey Epstein \(a\)](#) attempts to model Epstein's death and the organizations around it. However, see [this comment \(a\)](#) for pushback.

It's easy to make fun of Alex Jones tier conspiracy theories. But if we're being honest, it's really hard for any regular person to model opaque organizations like their local police department, their district attorney's office, the FBI, the NSA, the state department, or Congress. I think deep down most people are not conspiracy theorists, simply because they do not have the tools equipped to understand those organizations. Some of this is due to a lack of knowledge about what these organizations do and what their internal politics are. Some of this is due to the fact it's socially encouraged to have a non-sensibly cynical attitude when it comes to clandestine organizations, lest we be accused of being too naive by our wizened and grizzled friends.

But a lot of it is just because, by default, we no longer use the operationally important reasoning for understanding the behavior of people we actually know. Instead we feel free to shift into far-mode thinking, and posit relationships and arrangements that do not actually occur in the wild. The things our theories say about us and let us get to believe become more important than their predictive value. We don't actually see any of these grand coverups happen, but it's cool to imagine they do, especially when we get to imagine our political enemies doing it. Sometimes the long downtime between regime changes are so boring that it's easier and more exciting to just assume it's happening all the time, everywhere, right out of sight.

R Street, a libertarian policy think tank, offers an analysis of [A cybersecurity forecasting platform. \(a\)](#)

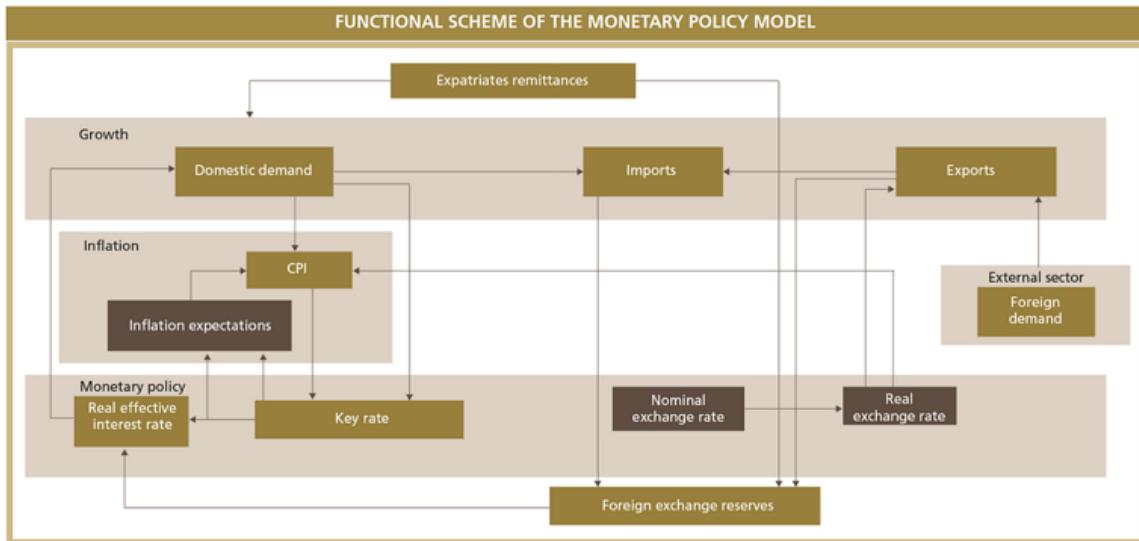
A new substack blogger and old Metaculus forecaster looks at [ordered prediction markets \(a\)](#), which might allow markets to extract probability distributions, rather than just probabilities. See also [here \(a\)](#) for a more academic treatment of the topic.

In the News

Google has revealed the existence of a new [internal prediction market \(a\)](#), with over 175,000 predictions from over 10,000 Google employees.

[NeuralProphet \(a\)](#) aims to be an update over [Facebook's Prophet library \(a\)](#), which offers time series data prediction in a box. I haven't tried it out, but it looks promising.

The Bank Al-Maghrib—the central bank of Morocco—has a very clearly written report on [monetary policy \(a\)](#). I found the international comparisons on growth, interest and inflation rates to be particularly interesting.



Some news media & individuals wrote some quantified predictions for 2022: [Vox \(a\)](#), [UnHerd \(a\)](#), [The Economist \(a\)](#), [Ipsos \(a\)](#), [Matt Rickard \(a\)](#), [Avraham Eisenberg \(a\)](#), [Mathew Yglesias \(a\)](#), [The Atlantic Council \(a\)](#), and [Blackrock \(a\)](#). h/t to Clay Graubard for [this longer list of 2022 predictions](#), from which some of the aforementioned were taken. It feels like there are more of these than last year, and the Mathew Yglesias piece is by a particularly mainstream author, which might be indicative that forecasting is becoming something less niche.

I tried to get Vox to make some bets against me [on](#) Twitter, but my megaphone wasn't big enough to get their attention.

Note to the future: All links are added automatically to the Internet Archive, using this [tool \(a\)](#). "(a)" for archived links was inspired by [Milan Griffes \(a\)](#), [Andrew Zuckerman \(a\)](#), and [Alexey Guzey \(a\)](#).

I critique here in such detail because, despite all our disagreements and my worries, I love and I care.

Zvi, [Thoughts on the Survival and Flourishing Fund \(a\)](#)

Forecasting Newsletter: Looking back at 2021

This is a linkpost for <https://forecasting.substack.com/p/looking-back-at-2021>

Table of contents

- The American Empire Has Alzheimer's.
- Prediction Markets: VC money, searching for DraftKings, predatory pricing, and the race to be last.
- For skilled forecasters, crypto prediction markets are much more profitable than forecasting platforms.
- Best forecasting pieces from 2021

You can sign up for or view this newsletter on [substack](#), where there are already a few thoughtful comments.

The American Empire has Alzheimer's

It is 1964. Sherman Kent is a senior intelligence analyst. While doing some rudimentary experiments, he realizes that [analysts themselves disagree about the degree of confidence that words convey](#). He suggests that analysts clearly state how certain they are of their conclusions, by using words that correspond to probabilities. This will allow keeping track of how analysts do, while being simple enough for less detail-focused politicians to understand. His proposal encounters deep resistance, and doesn't get implemented.

It is the 30th of April of 1975. With the [fall of Saigon](#), the US finally pulls out of a bloody war with Vietnam. There are embarrassing images of people flying out of the US embassy at the last moment. Biden is a newly-minted [senator from Delaware](#).



It is 2001. The US intelligence agencies are very embarrassed by not having been able to predict the September 11 attacks. The position of the [Director of National Intelligence](#), and an associated [Office of the Director of National Intelligence](#), is established to coordinate all intelligence agencies to do better in the future.

At the same time, [Robin Hanson](#) pushes for a "[Policy Analysis Market](#)", which would have covered topics of geopolitical interest. This proposal becomes too controversial, and gets dropped.

It is 2008. A bunch of Nobel Prize winners and other luminaries publish [a letter](#) urging the Commodity Futures Trading Commission (CFTC) to make prediction markets more legal.

It is 2010. IARPA, an intelligence agency modeled after DARPA, which incubates high-risk, high payoff projects, creates [a tournament](#) to find out which forecasting setups do best. Philip Tetlock had done some experiments which found that pre-selecting participants does pretty well. He repeatedly wins the IARPA tournament, and creates Good Judgment Inc to provide the services of his preselected high-performance forecasters. The US doesn't buy their services, and Good Judgment Inc survives by selling very expensive training sessions to clients which have too much money.

It is 2013. The CFTC shuts down [Intrade](#), one of the only prediction market platforms in the US.

It is 2017 and onwards. As Ethereum and crypto more generally become more mainstream, some prediction markets on top of crypto-blockchains start to pop up, such as Augur, Omen, and later, Polymarket. As cryptocurrencies become more and more popular, the fees on the original Ethereum blockchain increase so much that placing bets on these prediction markets becomes too expensive. Polymarket survives by moving to a "layer two" blockchain, a less paranoidly secure blockchain that mimics the Ethereum blockchain, and which allows users to continue betting.

It is the summer of 2021. Biden makes incredibly overconfident assertions about the Afghani government holding on against the Taliban. It doesn't. There are images of the evacuation from Kabul, Afghanistan which look very similar to the evacuation from Saigon, Vietnam. This is all very embarrassing to the Biden administration, and his approval rating drops drastically.

"There's going to be no circumstance where you see people being lifted off the roof of an embassy in the — of the United States from Afghanistan. [...] the likelihood there's going to be the Taliban overrunning everything and owning the whole country is highly unlikely." — Biden, [July 08, 2021](#)



Come Christmas of 2021, the CFTC gives Americans the gift of disappointment by shutting down Polymarket in the US, one of the few places where real money was being traded around topics of extreme interest to Americans, like US covid cases.

The picture I paint above is somewhat reductionist, and omits some important details. For instance, for a while the US had, and maybe still had an intelligence community prediction market. More recently, the United Kingdom also has a "Cosmic Bazaar", a forecasting tournament using [Cultivate Labs infrastructure](#). Dominic Cummings, who was Chief Adviser to Prime Minister Boris [talked of reading Scott Alexander's blogposts about Covid](#). There is also forecasting done with the Czech Republic, the Dutch and the OSCE. But I am yet to think that forecasters are meaningfully driving policy. The forecasting community is close-knit and I think there would be conversations if policymakers were regularly looking at forecasts—if you disagree, please get in touch.

Still, even after adjusting for my predisposition for pessimism, I think that the broad strokes of the above overview are about right. The US government is not being a "strong optimizer", whatever that means. In fact, the US government is being fucking dumb. But it took me a

while to crystallize this, and to notice how some of the dysfunctional aspects of the forecasting panorama have the same root cause. PredictIt's fees are so high (10%) because until very recently, they didn't have competition to keep them on their toes. Metaculus is—to some extent—structured around fake internet points because doing the real money version would have been bureaucratically exhausting.

Prediction Markets: VC money, searching for DraftKings, predatory pricing, and the race to be last

VC money

In the past few years, a few startups have joined the prediction market arena, chiefly Polymarket, Hedgehog Markets and Kalshi. Augur, previously a crypto project with very strong decentralization mechanisms, also sold out and spun off a more commercially oriented site, Augur Turbo, more focused on sports, crypto and entertainment. There were also a whole lot of less successful copycats, like [PolkaMarkets](#).

These projects have gotten a fair amount of funding. Polymarket got an initial \$4M investment round, and was [reportedly valued at \\$1B in later talks](#). Kalshi got \$30M in funding from, among others, [Sequoia Capital](#). Augur Turbo got an investment of [\\$1M from Polygon for its liquidity program](#), the network on which it and Polymarket runs. And Hedgehog Markets got a [\\$3.5M investment](#), as well as [\\$500k through the sale of NFTs](#), NFTs which allow users to participate in exclusive walled-off markets.

Searching for DraftKings

What is driving that valuation and initial investment? Well, for comparison, DraftKings, one of the biggest sports markets around, was valued at [\\$20B](#) before its stock price [took a beating after an adversarial report by Hindenburg's Research, a short-seller](#). So being a similarly large player in the nascent prediction markets field could be worth a significant fraction of that. Even being the "DraftKings of crypto", i.e.—the largest and more liquid player for sports within the crypto ecosystem—could potentially be worth quite a bit.

Predatory pricing and the race to be last

But first, these startups have to capture the market. And the way they are trying to do this is by subsidizing participation. That is, they create markets whose initial probabilities are off, giving users the chance to make money by participating in the market. I'm most familiar with how Polymarket has done this; I think they have overall lost money even though they have seen tens of millions in volume. My guess is that some other platforms have likewise lost a fair bit of money subsidizing volume.

But this creates a race to be the last to subsidize one's own markets, and steal the competition's user base. Then, perhaps, the last one standing could [monopolize the business and raise fees](#).

Alternative profit models.

In short, it looks that right now, there is money flushing around, but eventually the necessary sucker at the table will tend to be the users. And this incentivizes markets on sports, NFTs, or celebrities, rather than on war, politics, or technological developments, because they have more mainstream appeal. So the core, amoral business insight here is that all of these

platforms are vying for the same slice of the market: the sports and crypto markets. Or, in other words, entertainment for those newly rich off the crypto-boom.

What would an alternative business model be? Well, on the one hand, the different prediction markets could aim for different niches. Hedgehog markets could aim to entertain people heavily into the cryptocurrency scene. Polymarket could aim to be the best at real-world predictions. Augur could return to its original vision and be the go-to place for paranoid users interested in security. And FTX offers the best derivatives on cryptocurrency products.

Personally, the profit model that I'd like to see is one in which the prediction market platforms extract the profit not from their users, but rather from the people who are consuming the odds which the betting produces as a side-effect. For instance, a large NGO such as Open Philanthropy might be interested in a variety of geopolitical events, and could subsidize a market on them. This would involve providing both the liquidity (\$300 to a few thousand per market), and some money to support the prediction market platform. As the decentralized finance ecosystem develops, instead of a central organization paying for public goods, [DAOs](#) might form for this purpose.

Such a profit model would be such that the prediction market platform would be able to benefit in proportion to how much value it generates in the world. For instance, right now Polymarket creates value by producing common knowledge about sensible default probabilities to have around covid. Having sponsors which pay in proportion to how much value these markets produce, and which are willing to pay to create valuable markets might allow platforms such as Polymarket to capture a fraction of the value they create.

That kind of a profit model could, I think, make humanity more formidable.

For skilled forecasters, crypto prediction markets are much more profitable than forecasting platforms.

A drift divides the human forecasting space. On the first corner, we have forecasting platforms, which are legal throughout the land, and which see lower volumes. Forecasters play by giving their probabilities, and checking whether these are more right than other participants'. They are often rewarded according to a proper scoring rule so that they're incentivized to give their true and honest best guess, but this often fails in [amusing ways](#). Chief amongst these platforms are Metaculus and Good Judgment Open, which have many questions and whose communities have historically been open and welcoming. Forecasting platforms tend to have socially useful questions on e.g., geopolitics, technology developments, or risks to society.

On the opposing corner, we have prediction markets, where participants put their money where their mouth is, and earn money if they turn out to be right. The communities can also be welcoming in their own ways, though this is partially because good bettors are looking for less experienced bettors to fleece. Prediction markets are of dubious legality, mostly because some apparatchiks under the Commodity and Futures Trading Commission (CFTC) have a hard-on against it. In principle, real-money prediction markets could have questions on any topic, but they tend to have questions that have more mainstream appeal, such as on sports.

In recent times, it has become noticeably the case that prediction markets, and in particular crypto prediction markets, are significantly more profitable to participants than forecasting platforms. So some forecasters who trained themselves on Metaculus then flocked to try their luck on Polymarket, and the best ones made significantly more money than they could have made on Metaculus.

For reference, the current monetary rewards given in forecasting platforms are roughly as follows:

- Metaculus: On the order of \$1000 per tournament.
- Hypermind: On the order of \$5000 per tournament.
- Replication Markets: Around \$150k in total, for around 10-20 rounds of forecasting, each of which contained many questions.
- CSET-Foretell; \$120k a year, or \$200 per forecaster per month.

Note that these are per tournament, so if a \$1000 Metaculus tournament contains around ten questions, and ten forecasters participate on it, this amounts to \$10 per forecaster. If a forecaster spends more than an hour per question, they are then earning less than minimum wage.

In contrast, prediction markets, such as PredictIt or Polymarket often see upwards of \$100k traded on individual markets. So the top predictors can and do make a comfortable living betting, in a way that would be difficult or impossible to do in forecasting platforms rather than in prediction markets.

For instance, one of the top earners on [Replication Markets](#) earned around \$10k. But he spent significant time programming tools to automate his trading, and it seems like he put a lot of love into his forecasting. If he had invested that labor into trading on Polymarket and building tools to do so, or if he had simply sold his labor as a programmer, he would have earned significantly more money.

So the incentives are not pointing in the right direction. Capable forecasters can earn significantly more by predicting societally-useless sports stuff, or simply by arbitraging between the big European sports-houses and crypto markets. Meanwhile, the people who remain forecasting socially useful stuff on Metaculus, like whether [Russia will invade the Ukraine](#) or whether there will be any [new nuclear explosions in wartime](#), do so to a large extent out of the goodness of their heart.

I think that the clear solution to this is to either increase the overall willingness to pay forecasters, or to be willing to subsidize liquidity in prediction markets for questions that are of general value.

Best pieces on forecasting during 2021

Practice

[Predicting Politics](#) is generally worth reading, starting with [How to get good](#), [Mining the Silver Lining of the Trump Presidency](#), and [Boring is back, baby](#).

Avraham Eisenberg wrote [Tales from Prediction Markets](#), gathering a few interesting anecdotes.

[Cultured meat predictions were overly optimistic](#) (a). "Overall, the state of these predictions suggest very systematic overconfidence."

Charles Dillon of Rethink Priorities and SimonM looked at [How does forecast quantity impact forecast quality on Metaculus?](#) More forecasters increase forecast quality, but the effect is small beyond 10 or so forecasters.

David Friedman looked at whether the [past IPCC temperature projections/predictions have been accurate?](#)

[Violating the EMH — Prediction Markets](#) gave specific examples in which prediction markets appeared to violate the efficient market hypothesis.

[How I Made \\$10k Predicting Which Studies Will Replicate](#). The author started out with a simple quantitative model based on [Altmejd et al. \(2019\)](#), and went on from there.

Together with my coauthors Misha Yagudin and Eli Lifland, I posted a fairly thorough investigation into [Prediction Markets in The Corporate Setting](#). The academic consensus seems to overstate their benefits and promisingness. Lack of good tech, the difficulty of writing good and informative questions, and social disruptiveness are likely to be among the reasons contributing to their failure. In the end, our report recommended not having company-internal prediction markets.

Charles Dillon wrote [Data on forecasting accuracy across different time horizons and levels of forecaster experience](#), using Metaculus and PredictionBook data, and building on earlier work by [niplav](#).

Futurism

[Incentivizing forecasting via social media](#) explored the implications of integrating forecasting functionality with social media platforms.

The Machine Intelligence Research Institute's research on agent foundations shed some light on [probability theory more generally](#), [Radical Probabilism](#) and [Reflective Bayesianism](#) seem particularly worth highlighting, as does [Probability theory and logical induction as lenses](#).

Daniel Kokotajlo wrote [What 2026 looks like \(Daniel's Median Future\)](#), extrapolating the performance of models like GPT-3 year by year. Ben Snodin wrote [My attempt to think about AI timelines](#).

The Machine Intelligence Research Institute has published a few [conversations on future AI capabilities](#). Of these, readers of this newsletter might be particularly interested in the [Conversation on technology forecasting and gradualism](#).

Theory

There was some back and forth online on Kelly betting:

- [Kelly isn't just about logarithmic utility](#)
- [Kelly is just about logarithmic utility](#)
- [Never Go Full Kelly](#)
- [Why the Kelly criterion kind of sucks](#)

See also: [Proebsting's paradox \(a\)](#), a thought experiment in which naïve Kelly bettors are lead to ruin, [Learning Performance of Prediction Markets with Kelly Betting](#) proves that prediction markets with Kelly bettors update similarly to Bayes' law, and [this blog post](#) illustrates that paper's point in a more approachable manner.

After reading these posts, I'm left with the conclusion that Kelly betting is an interesting yet ultimately limited tool. One encounters the limits of applicability as soon as one is exposed to many bets at once, or to the chance that the bets may change in favorable or unfavorable directions.

Alex Lawsen and I published [Alignment Problems With Current Forecasting Platforms](#), outlining problems with the incentive mechanisms in almost all non-prediction market platforms.

[The Generalized Product Rule](#) outlines how a certain step in [Cox's theorem](#)—the step which proves that probability updating is multiplicative—can be applied to other problems as well.

Jaime Sevilla took a deep dive into [aggregating forecasts](#).

Eli Lifland published an article [on bottlenecks to more impactful forecasting](#). It crystallizes his knowledge from a few years of his forecasting on Metaculus, CSET-Foretell and Good Judgment Open.

Forecasting Newsletter: January 2022

Highlights

- Polymarket now operates out of [Panama](#)
- Aver, a new Solana-based betting site, raised [\\$7.5M in seed funding](#)
- Announcing the Forecasting Newsletter \$10,000 Micro-Grants Program
- Forecasting meetup next Wednesday (9th Feb) at 19:00 UTC in the [LessWrong online garden](#)

Index

- Prediction Markets & Forecasting Platforms
- Forecasting Newsletter Micro-Grants Program
- Blog Posts

You can sign up for this newsletter on [substack](#), or browse past newsletters [here](#). If you have a content suggestion or want to reach out, you can leave a comment or find me on [Twitter](#).

Prediction Markets & Forecasting Platforms

Polymarket

Since Polymarket [settled with the CFTC last month](#) (a), they have moved out of US jurisdiction. They are now operating from [Panama](#) (a). Their terms of use now prohibit users from the US from participating on the platform, and likewise [prohibits use of a VPN to circumvent this prohibition](#) (a). They have now started using [UMA](#) (a) as a market resolver. Personally, I am glad Polymarket has survived.

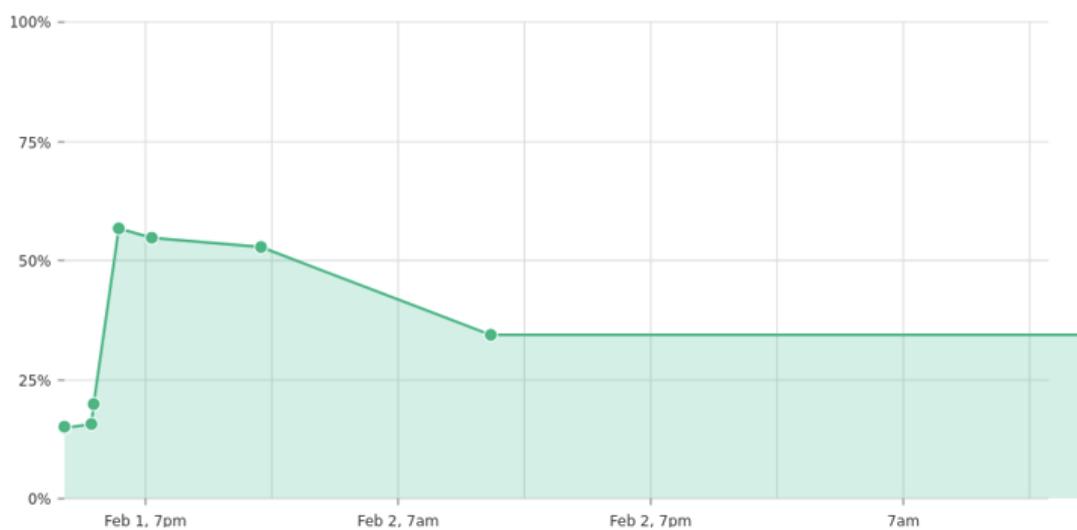
Manifold Markets

[Manifold Markets](#) (a), a new play-money prediction market, has kept its frantic development pace. They will be "launching" on February the 8th, from their current open beta. They outline some of their hopes and plans in their [substack newsletter](#) (a).

Will I find that the PIBBSS Fellowship was a success?

34% chance

Nuño Sempere • Feb 1 - Feb 8, 2022 • M\$ 262 pool



I created some markets on whether I will consider the [Principles of Intelligent Behavior in Biological and Social Systems](#) fellowship to be "[a success](#)" (a), on [the number of subscribers to this newsletter](#) (a), and on whether [my current employer will still be alive by the end of 2022](#) (a). The process was extremely painless, and I recommend that readers [give it a try](#).

CSET-Foretell is now INFER

CSET-Foretell moved from being hosted by [CSET](#) (a) at Georgetown to being hosted by [ARLIS](#) (a) at Maryland. ARLIS also received an [\\$8M grant](#) (a) from Open Philanthropy for forecasting specifically.

This is a fairly large amount of [EA](#) funding for forecasting projects. For reference, Metaculus recently received [\\$250k](#) (a), and Phil Tetlock and associates previously received [two](#) (a) [~\\$500k](#) (a) grants. Open Philanthropy also previously gave ~\$55M [CSET](#) (a), as well as later [top-up](#) (a) [grants](#) (a). But this was for a broader think tank, rather than for forecasting specifically.

In [the December issue](#) of this newsletter, I assessed that the move from CSET to ARLIS was probably a negative development, partially because I thought that funding from Open Philanthropy was much better than government funding. As it happens, ARLIS has just now received funding from Open Philanthropy as well. Multiple people also reached out to comment that the move was probably neutral or positive, on account of ARLIS' deeper involvement with the US government. My independent impression is that I still dislike the move, but my all-things-considered view is that it's probably ok and I was wrong. As to how wrong, we'll see.

Rank	Team	Scored Questions	Overall Score
1.	Samotsvety Forecasting	24	-1.105
2.	Ariana's Team	24	-0.586
3.	Olavo's Team	13	-0.556
4.	Aisha's Team	14	-0.49
5.	Pro Forecasters	16	-0.477
6.	Wake Forest University	6	-0.384
7.	Saaima's Team	13	-0.178
8.	Shaun's Team	24	-0.166
9.	Mariana's Team	4	0.046
10.	Molly's Team	12	0.076

Separately, because their Pro Forecaster program still only pays \$20/hour, my team—Samotsvety Forecasting, which overwhelmingly won the last two seasons—might not be participating going forward, though we are trying to negotiate with them. I talked to a few super-forecasters about this, and \$20/hour isn't going to get ARLIS the best forecasters. Their open call for pro forecasters can be found [here](#) (a).

But this isn't a problem unique to INFER. Generally, forecasting platforms such as [Metaculus](#) (a), [Hypermind](#) (a), or the [Social Science Prediction Platform](#) (a), just don't pay that much to forecasters. This leads to predictable problems, such as forecasters [moving to crypto prediction markets](#), or experiencing burnout after forecasting in an unsustainable way for little reward, as happened to some top covid forecasters.

Metaculus

[SimonM](#) (a) kindly curated the [top comments from Metaculus this past January](#) (a). They are:

On the potential invasion of Ukraine by Russia:

- [Tilter](#) (a) believes that Russia's slow deployment of troops is a move intended to scare NATO into keeping out of Ukraine. If Russia was going to invade, they would just do that quickly, as has been their manner of operating in the past.
 - [jmason](#) (a) points out that the US State Department just issued a Do Not Travel warning for Ukraine and [Belarus](#) where Russian forces are massing for a declared joint-military exercise.
 - Note that most of the major platforms have forecasts on Russia/Ukraine as well. They can be found on [Metaforecast](#).
- [EvanHarper](#) (a) emails NASA to resolve a Metaculus question.
- [alwaysrinse](#) (a) notices that Scott Alexander successfully trolled the Metaculus community.

Metaculus' updated public figure predictions can be found [here](#) (a).

Forecasting meetup

[Nathan Young](#) and I are organizing a forecasting and prediction markets online meetup at 7:00 PM UTC on the [9th of February in the LessWrong online Walled Garden](#). The LessWrong Walled Garden is great because you can move around and leave and join conversations as you wish, allowing better flowing conversations, mentions Nathan. The event will officially finish 2 hours after it starts, but anyone is welcome to stay for longer.

Odds and ends

The sixth [Makridakis Competition](#) (a)—a well-known open competition to evaluate and compare the accuracy of different time series forecasting methods—has been announced. Total prizes—amongst many rounds and many forecasters—[sum up to \\$300,000](#) (a). The deadline for the first submission point is the 6th of March, though there is a trial round in February. More details can be found on their [webpage](#) (a)

[Aver](#) (a), a prediction market built on top of the Solana chain, announced a [\\$7.5M seed round](#) (a). My impression is that they will start with sports and crypto, which are easier to resolve. But I imagine that they will also have the flexibility to also experiment with some markets whose probabilities are valuable to the general public. Based on the typical amount sold during the first financing round, their valuation is probably between \$30M and \$80M.

Insight Prediction has launched a real-money beta with limited access and is looking for members of the prediction market community to test the site and give feedback. They are running a [Russian Invasion of Ukraine Market](#) (a), as well as a weekly [US covid total death markets](#) (a). Their rationale is that these might make some people uncomfortable, but they think that these markets are the most important for policymakers and users in terms of the information they provide. I agree with that reasoning.

On the negative side, Insight Prediction had previously been stuck in development for a long time. They were originally planning to launch in June or July 2021, though at the time one of the funders refused to take a \$20 bet I offered on their timelines. Reliable anonymous sources have also expressed some skepticism about the project—not necessarily in the sense of being a scam, but rather in terms of their plans being unfocused.

Will Russia Invade Ukraine in 2022?

Bid price	Bid size	Ask price	Ask size
14.1	100	22.5	25000
14	250	24.9	300
11	3000	25	100
7	100	33	100
1	5000	99	5000

Maximum Buy Price: 22.5 Number Of Shares: 25000

The currency of this market is USD. 100 cents = 1 USD.

Buy Sell

Next

Skip Offer Confirmation

Dates and Currency

Market Open Date	Market Close Date	Currency
2022-01-20	2023-11-01	USD

Prediction market players who want to participate in the early access beta can reach out per [email](#).

Announcing the Forecasting Newsletter \$10,000 Micro-Grant Program

After the apparent success of [ACX grants](#) (a), [Misha Yagudin](#) and I received \$10k from an anonymous donor to give out as micro-grants through this newsletter.

Some examples of projects we'd be excited to fund might be:

- Open-source software and tooling to automate forecasting. Bonus points if Metaculus users or other forecasters start using it.
- Giving a shot at forecasting or making models of a difficult yet useful and decision-relevant area. Think of [Rootclaim](#) (a) analyzing the lab-escape story of Omicron.
- Pieces similar in quality to the ones mentioned in "[best pieces on forecasting from 2021](#) (a)", in [Forecasting Prize Results](#) (a), or in [some possible research areas](#) (a).
- Trying to estimate many uncertain parameters, e.g., the quality of all US or UN organizations, quality of academic fields, whether a large list of organizations will fail, enlightened willingness to pay for many products, the accuracy of many public figures, etc.
- Create a [microcovid](#) (a) or [foodimpacts](#) (a) but for other areas, like micro-marriages, micro-insights, micro-dooms, etc. Do this in a way that easily allows the creation of many of these calculators.
- Improve [metaforecast](#) (a) (which is [open source](#) (a)) in some interesting way, e.g., improve the estimates of forecast quality.

The application form is [HERE](#) (a). Feel free to apply for more than \$10k: we don't anticipate having much difficulty getting more funding for promising applications, and we may refer these to other funders (e.g., the [EA Infrastructure Fund](#) (a)) if we can't.

We will be accepting applications until March 15, though we may extend this period if we don't receive enough high-quality submissions. We preliminarily plan to make decisions by April the 1st.

Otherwise, Luke Muehlhauser [comments](#) (a) that forecasting related projects might be a good fit for the [EA Infrastructure Fund](#) (a). Jonas Vollmer, who runs EA Funds, [confirms this](#) (a)

For larger projects, the Survival and Flourishing Fund, backed by philanthropists Jaan Tallinn and Jed McCaleb, is organizing the distribution of around [\\$6M-\\$10M in grants](#) (a) this June, with applications due on Feb 21. They generally only accept applications from registered charities, but [speculation grants](#) (a) might be a good fit for smaller projects (40%).

Blog Posts

Scott Alexander posts [Predictions for 2022](#) (a) and [grades his own probabilities for 2021](#) (a). Zvi [gives his own probabilities for them](#). (a).

Ege Erdil writes about [retrospective forecasting](#) (a): In order to use Bayes' theorem, we need to somehow get the probabilities before the thing we are interested in happened. He links to [this interesting blogpost/paper](#) (a) extracting a 42% that the Confederacy would win the US civil war from the value of their gold bonds in Amsterdam. On this topic, see also Lustick and Tetlock's [The simulation manifesto](#).

Note to the future: All links are added automatically to the Internet Archive, using this [tool](#) (a). "(a)" for archived links was inspired by [Milan Griffes](#) (a), [Andrew Zuckerman](#) (a), and [Alexey Guzey](#) (a).

h/t to [Nathan Young](#), [Ozzie Gooen](#) and [Clay Graubard](#) for comments and suggestions on this edition.

Jefferson, in his forecast, had anticipated this, as the "rock upon which the old Union would split." He was right. What was conjecture with him, is now a realized fact.

Alexander H. Stephens, [Cornerstone Speech](#).

Forecasting Newsletter: February 2022

Highlights

- The FTX foundation will potentially give out millions in forecasting
- Insight Predictions and Futuur have real-money prediction markets on the invasion of Ukraine
- At least \$444k paid out in over the counter bets on the invasion of Ukraine

Index

- Thoughts on The FTX Foundation Funneling Funds to Forecasting
- Prediction Markets & Forecasting Platforms
- Forecasting Job Board
- Odds and Ends
- Long Content

You can sign up for this newsletter on [substack](#), or browse past newsletters [here](#). If you have a content suggestion or want to reach out, you can leave a comment or find me on [Twitter](#).

Thoughts on the FTX Foundation Funneling Funds to Forecasting

The FTX Foundation announced a massive [\\$100M to \\$1B/year \(a\)](#) Future Fund. Amongst their [areas of interest \(a\)](#) and [project ideas \(a\)](#) are:

More forecasting. We're huge fans of prediction markets and forecasting tournaments. We'd love to see these widely adopted and used to inform political decision-making. We're particularly excited about long-term forecasting (10+ years out), and methods that might make long-term forecasting more feasible.

Prediction markets. We're excited about new prediction market platforms that can acquire regulatory approval and widespread usage. We're especially keen if these platforms include key questions relevant to our priority areas, such as questions about the future trajectory of AI development.

Forecasting Our World in Data. We'd love to see a project that takes one hundred of the most important charts in Our World in Data (we think the Technological Progress charts would be especially interesting) and employs superforecasters to plot out how the charts will go over the next one, three, ten, thirty and one hundred years. Ideally, the output would be well-presented and easily understandable, and display probability distributions for each year.

Forecasting that will affect important decisions. We think a key challenge for making forecasting organizations better is ensuring that the questions asked are interesting and important. We'd be especially excited about forecasting projects that have a great plan for ensuring that the questions asked are of significant interest to influential and altruistic actors, potentially including thoughtful government officials and large funders in the EA ecosystem.

More generally, we're interested in a "superforecasting institute." Few jobs are more important than rigorously forecasting the future, but currently it's hard to do that job full-time. We want to allow excellent forecasters to make superforecasting their career. And we

want to explore creating prizes and fellowships that will optimally incentivize outstanding forecasting work.

They also have a [project ideas competition](#) (a), which closes by Monday the 7th, which feels too short, as well as various other [applications](#) (a) on their website.

In comparison and contrast to Open Philanthropy, they seem to be moving fairly quickly. I'm hoping they will donate to smaller and nimbler forecasting projects which have a chance to be very valuable, rather than to larger, already established projects that are more sure to produce a perhaps more certain but also perhaps more mediocre impact.

Some signposts to look at will be:

- Whether they will use crypto-prediction markets (+)
- Whether they give money to Tetlock collaborators (-)
- Whether the projects they invest in will pay forecasters well (+)
- Whether they invest in projects that interact with large bureaucracies (--)
- Whether they will use forecasting tournaments (\dot{z} -?) or real-money prediction markets ($\dot{z}+?$)
- Whether forecasting systems will be used to legitimize what one believes (-), or to find out what is the case (+)
- Whether they will use Hypermind, CultivateLabs infrastructure (-), or more innovative platforms, like Manifold Markets (+)

The observation that I'm trying to point at is that there are sure options that [have been tried before](#) (a) and [mostly failed](#) (a), and innovative options which might feel more risky, and might yet fail, but which explore uncharted lands.

That is, there is a spectrum between three letter intelligence agencies and Polymarket, between superforecasters-trademark-registered and [rogueish crypto traders](#) (a), between the [\\$2,500.00 Keep Virginia Safe Tournament](#) (a) and the [Russian Invasion of Ukraine question](#).

My sympathies lie with the later. There are tradeoffs between exploration and stability, between moving fast and being really legible to outsiders, or between interacting with large bureaucracies and everything else. And because forecasting is not yet as useful as I think it could be, I mostly think that exploration is the right choice.

This is not to say that large projects that interact with large bureaucracies such as the US government using Cultivate Labs' stable infrastructure don't have a place. In particular, the tried and true options allow one to conserve [weirdness points](#) (a) while doing something weird somewhere else.

But with FTX spending so much money, they will also get to shape the forecasting community as a whole. Ideally, I would prefer to see a [fully alternative stack](#) (a) in which people can radically focus on "doing the thing". FTX's Fund would be in a position to implement such a thing. But they probably won't. Still, I'm curious about whether FTX's Fund will be willing or able to identify and fund tasteful and ambitious forecasting projects while moving so much money, so fast.

Prediction Markets & Forecasting Platforms

Metaculus

[SimonM](#) (a) kindly curated the [top comments from Metaculus this past February](#) (a). They are:

- [New highest voted comment of all time](#) reports how Metaculus helped an Ukrainian escape Kyiv. "Just want to say that I moved from Kyiv to Lviv on Feb 13 /entirely/ thanks

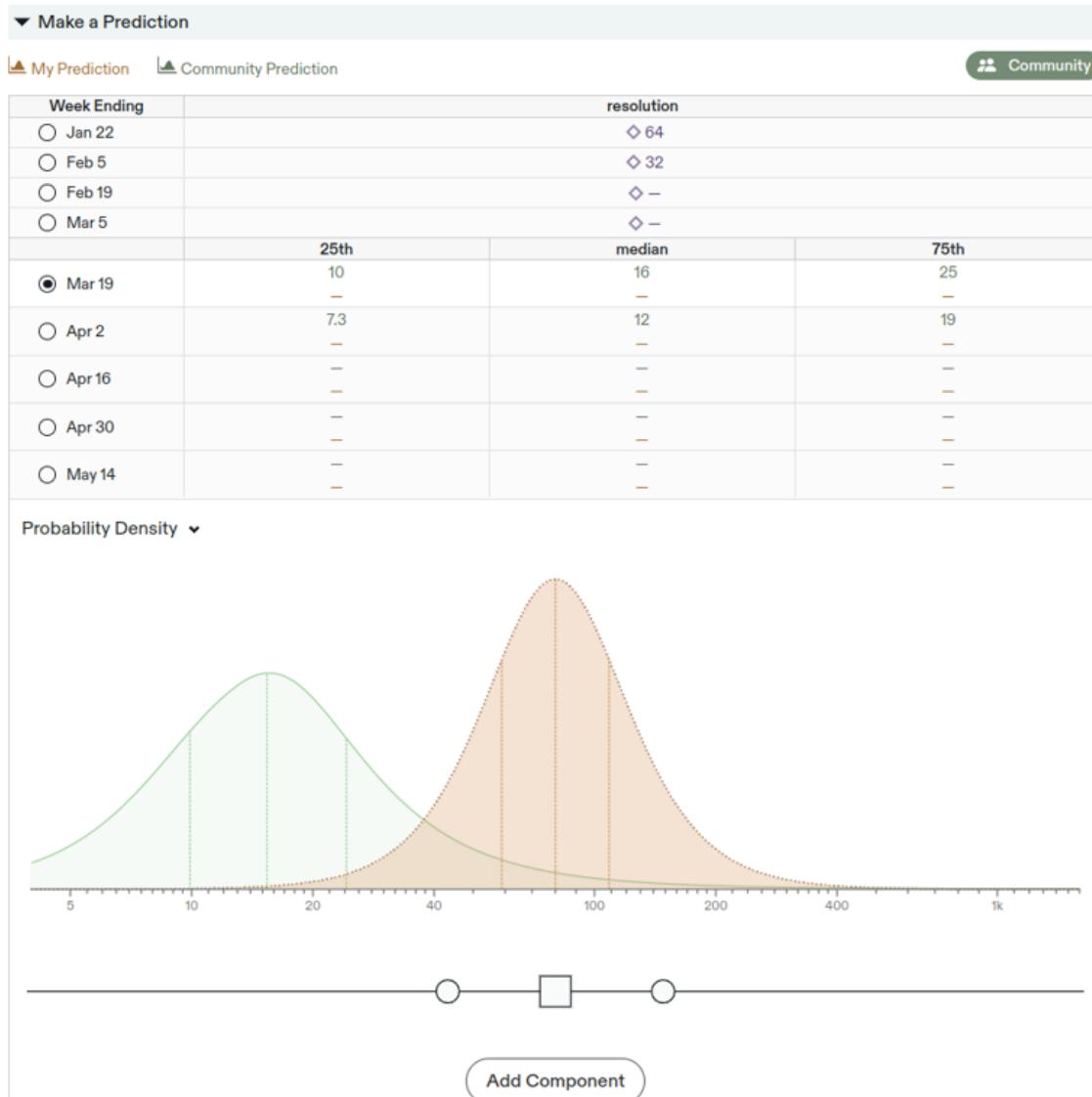
to this prediction thread and the Metaculus estimates. (Still in Lviv but leaving Ukraine later today.)."

- [Russia invades Ukraine](#)
- User [Laplace](#) ([a](#)) calculates the base rate of Russia invading its neighbours, using a Poisson process.
- fianxu calls question writers [to arms](#)
- wobblybobby looks at the [distribution for casualties in the Ukrainian-Russian war](#).
"Bottom line, estimating under 25,000 deaths requires a near term end to the war. If it drags on, deaths in excess of 50,000 to 100,000 are very likely."

Due to the war in Ukraine, there were an increased number of [quality contributions](#) ([a](#)), which might be worth reading. Metaculus mobilized more money to pay for forecasters and created many on-topic questions on short notice. Kudos!

In particular, they created the [Ukraine conflict tournament](#) ([a](#)), which already has 60 questions and a \$10k prize pool. They also quickly recruited forecasters and put some thought into how to best structure high-risk forecasting. For instance, some nuclear questions have gone private. Forecasters who already predicted those questions can still see them, so some of you might not have noticed. Metaculus is reviewing their policy on questions like these; they're working hard and will update when they're ready. If you would like specific questions you can submit them as normal, or send a message to [Nathan Young](#) ([a](#)), who has been writing questions around this.

Metaculus also added a feature allowing forecasters to predict on the same question at different points in time. So far, it seems to only be available on [questions](#) ([a](#)) in the [Flu Sight](#) ([a](#)) tournament.



Polymarket

Polymarket used the more novel [Uniswap 3](#) (a) algorithm to provide liquidity during the [Super Bowl](#) (a). This allowed users to bet larger amounts without the odds moving as much.

They also introduced a [liquidity mining/trading rewards program](#) (a), to subsidize participants to add liquidity (automatic market-making), as well as high-volume traders. The hope in the community is that this could avoid some damaging kinds of front-running bots ([sandwichers](#) (a)) by increasing fees for everyone and then doing rebates to all users except those that take part in malicious behaviour.

Polymarket is currently using [UMA](#) (a) to resolve their markets. As explained on the Polymarket Discord by Monsieur Dimanche, a well-known community member (taken with permission, and lightly edited):

There are no markets resolved by the Polymarket team anymore. Everything goes through UMA. However, Polymarket still needs to ask UMA to resolve markets, and users can still

use the Discord channel to tell Polymarket that we think a market has met the criteria for resolution.

What happens next is this: Poly thinks a market is ripe for resolution, so they ask UMA for a settlement. To incentivize this settlement, they give UMA a small amount of money (think 10 to 50 USD) that will be used as rewards. Once this is done, the market lands on oracle.umaproject.org (a).

At this point, anyone can propose an outcome for the market in order to earn the reward that Poly gave UMA. But when you do it, you have to bond a large amount of money (thousands of dollars). So you have to be careful before submitting an outcome: if you submit a wrong outcome that is rightfully contested, you will lose your bond. Of course this is intended behaviour, meant to heavily discourage people from proposing bad answers.

Here's an example. The market for "below 100k cases before April 15" has a current proposed answer of "yes" (meaning it did in fact happen before April 15). If you click on the market you get to this page:

The screenshot shows a web-based application for a binary market. At the top, there are tabs for 'POLYGON' and 'PROPOSAL'. Below the tabs, the title 'YES_OR_NO_QUERY' is displayed, along with a long hex string 'Request: 0x4bedd3319d0b799549cd09e08550fb...'. On the left, a 'Dispute Period' section contains a text input field with '1.0 = yes' and a red 'Connect wallet' button. Below this is a proposal ID: 'Proposal: 0xfc1d7f908c7fd97f5db21fd2beader10019853ae4806e8459c955411a0a1563a'. On the right, a 'Parameters' section lists: Disputer bond (USDC 11,500), Disputer reward (USDC 10), Sponsor Reward (USDC 10), and Liveness period (Time remaining: 00 h 42 min 38 sec left). There is also a warning icon with a red exclamation mark.

If nobody contests the proposed answer, the market will resolve "yes" in 42 minutes. Anyone can dispute an outcome but as you can see, it costs \$11500 to contest, and of course you lose that amount if you're wrong (if you're right however, you get it back, and it's the original proposer who loses the 11k). This is quite expensive and should deter people from trying dumb contests, like the ones that plagued Augur during the 2020 election aftermath.

Should an answer be contested, the price to contest would escalate and in the end go to a vote where all UMA tokenholders could vote on the correct outcome.

In comparison, the first contests for Augur were done with a tiny amount. I remember traders being annoyed that you could block the resolution of a multi million dollar market with the price of a movie ticket.

Another difference is that the vote would happen within 2-4 days, whereas the Augur process was painfully slow. This, coupled with the high threshold for contest, makes it a vast improvement in practice, even if the general idea stays the same.

On account of reading this, I bought a medium amount of the UMA governance token on [Uniswap](#) (a). Polymarket previously was a few steps ahead of the competition when choosing Polygon, and if they are displaying a similar degree of foresight when choosing UMA, the price of its governance token could likewise go up. Also, "a version of Augur that actually works" is a pretty enticing proposition. This is not investment advice, etc., etc.

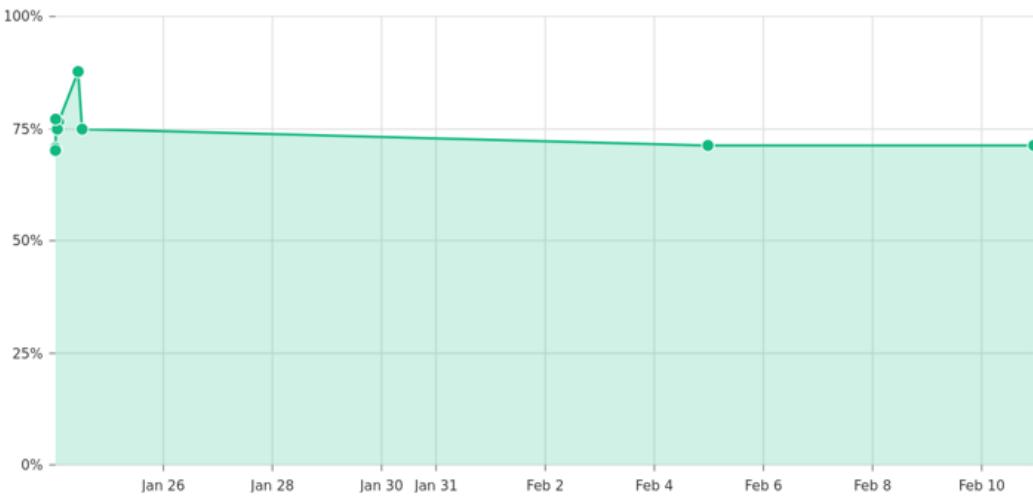
Manifold Markets

Manifold markets [received an EA grant](#) (a)

Will Manifold Markets win an EA Grant?

Resolved
YES

 Austin  Jan 24 - Feb 11  M\$ 103 pool  Tweet

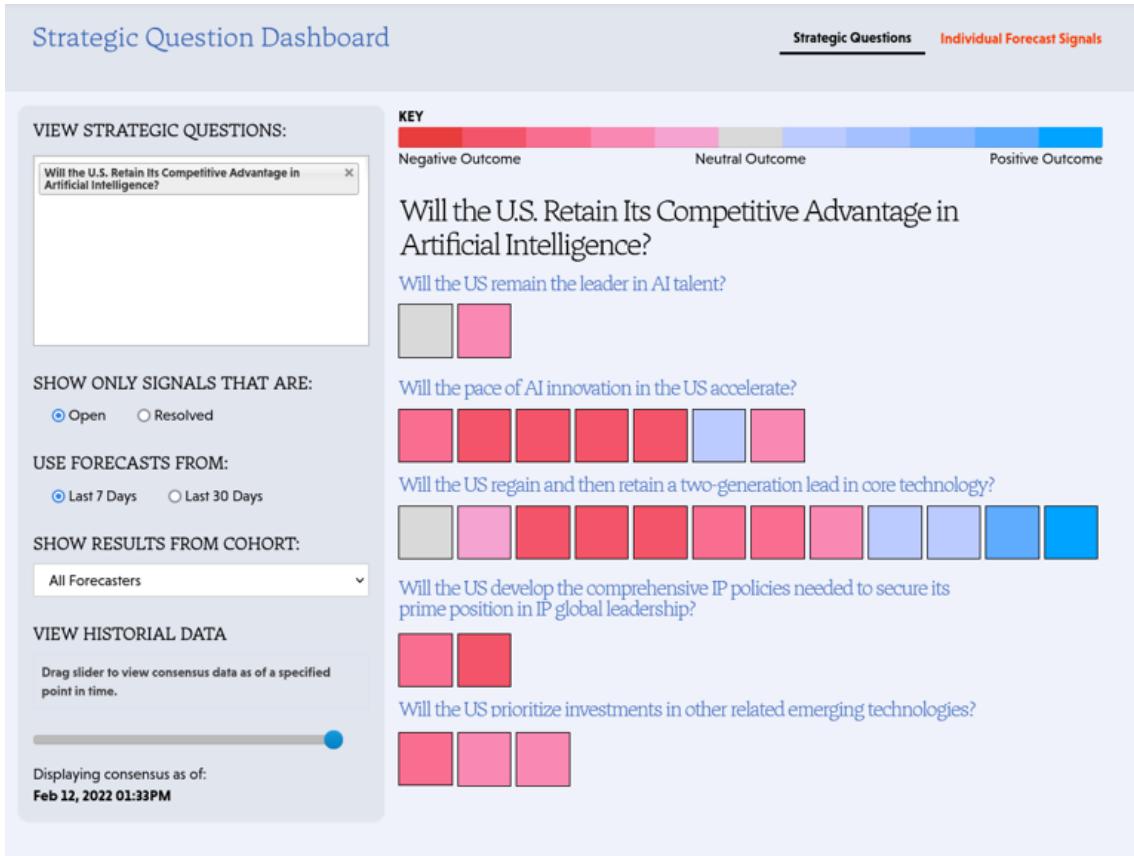


They report over their updates at [Above the fold](#) (a): they've been adding new features at a steadily fast pace. For instance, Manifold now supports free-form answers. So when betting on the 2024 election, one could have an initial lineup including the expected candidates, but if a dark horse candidate rises to prominence, it could later be added.

Manifold also released a [beautifully documented API](#) (a).

INFER

INFER released a [few blogposts](#) (a) outlining their current thinking and future plans. Of these, [Understanding strategic question decomposition](#) (a) is worth reading as a cute illustrated recap of the [best current approach](#) (a) for using forecasting systems to give insight on big picture questions.



They are also running a lottery to give \$2,000 to one lucky forecasting team. Teams have to be of 6 people, and the lottery is such that chances are maximized if they predict every day. Suppose that making a forecast one is not ashamed of takes 5 minutes and that 5 new teams are created. Then the expected prize winnings per hour are $\$2000 * 60 \text{ mins per hour} / (5 \text{ teams} * 5 \text{ mins per forecast per day} * 30 \text{ days} * 5 \text{ forecasters per team}) = \$26 / \text{hour}$, or not enough for me to do it.

INFER also tweaked its algorithm for aggregating predictions to give more deference to better forecasters.

Forecasting Job Board

Cultivate Labs, the company that maintains the forecasting infrastructure behind Good Judgment Open, INFER, and the Cosmic Bazaar, is [hiring \(a\)](#) for Government Consultant and Senior Rails Developer positions. Applicants must be US citizens.

Amazon is hiring for [Senior Program Manager, Network Forecasting and Planning \(a\)](#), as well as for an ["Applied Scientist" \(a\)](#) role for one of their forecasting teams.

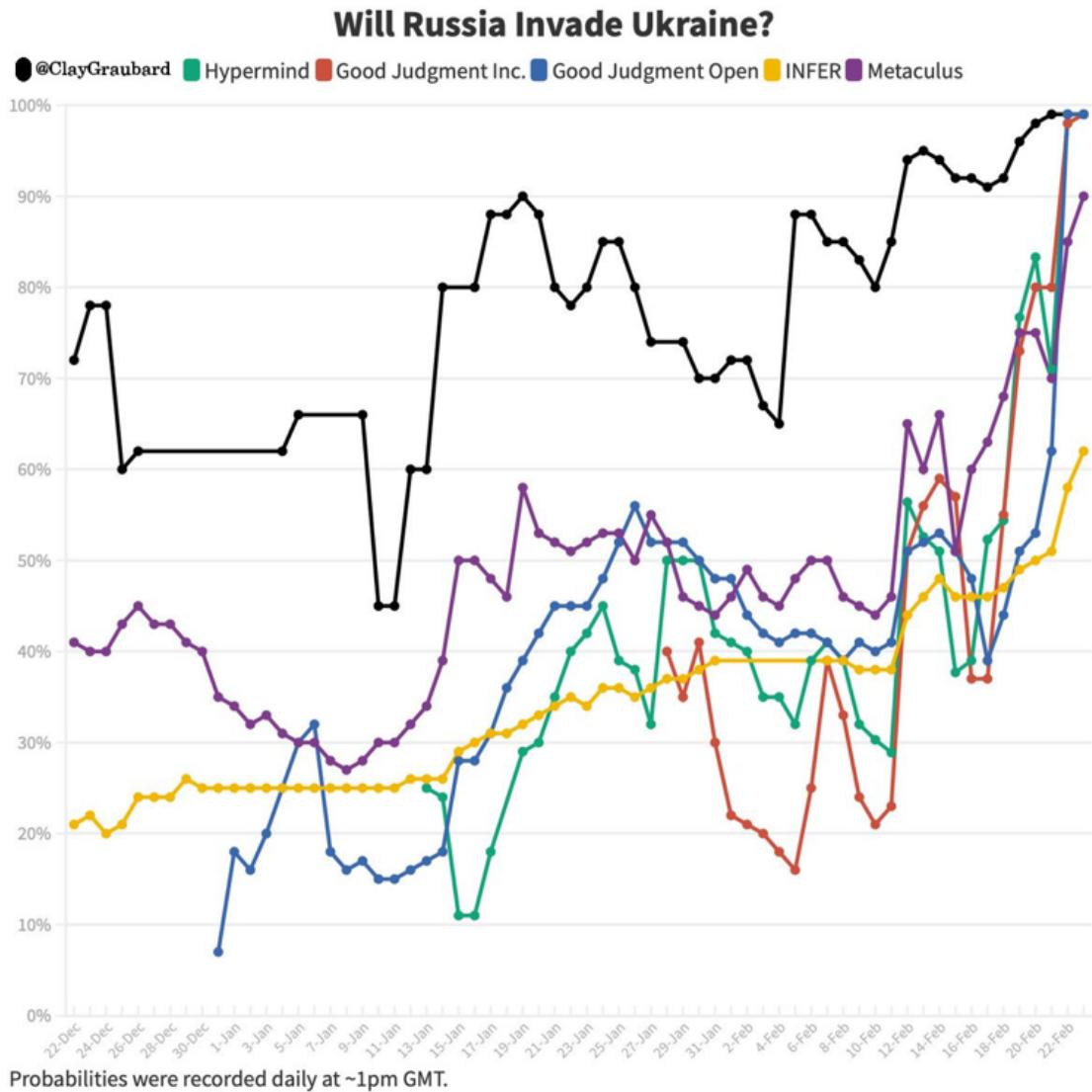
The [Quantified Uncertainty Research Institute \(a\)](#), the non-profit for which I work, will be hiring for researcher, software engineering, operations specialists, and product manager positions. We write software (like [Squiggle \(a\)](#) or [Metaforecast \(a\)](#)), and write [research \(a\)](#). If that sounds interesting, consider [reaching out](#).

Separately, my network currently has more opportunities for forecasting consulting, tournament creation, and general forecasting-related work than we know what to do with them. If you are an excellent forecaster or an excellent organizer, consider [reaching out](#).

Finally, an anonymous benefactor increased the size of this newsletter's [microgrants program](#) ([a](#)), so if you have a forecasting or epistemics-related project you'd be keen to implement, consider applying. We recently gave our first \$5k grant to Clay Graubard, for work related to his [quantified journalism](#) ([a](#)) on the Ukraine invasion.

Odds and Ends

Clay Graubard collects how the different forecasting platforms did at predicting the invasion of Ukraine. He [describes the situation](#) ([a](#)) as "not the forecasting community's finest hour". It's not clear to me that this is a fair assessment:



Not pictured there are prediction markets such as [Insight Markets](#) ([a](#)), where my forecasting group and I won \$20k betting on the Russian invasion, or [Futuur](#) ([a](#)), which likewise has real money markets on Ukraine.

Although I'm fairly sure they're not, they could yet be scams, so prospective participants should tread carefully. That said, I admire the courage of these two platforms for having markets on this topic.

The forecasting community also saw a few over-the-counter bets on Ukraine:



iabvek @iabvek · Feb 18

...

bet my \$2.5k to [@TarasBob](#)'s 7.5k that russia invades ukraine (resolved the same way as this metaculus market)

metaculus.com/questions/8898...

(hint to metaculus users -- your fake money prices should not be >40c away from the equilibrium price real money bets are being made at)



metaculus.com

Will Russia invade Ukrainian territory before 2023?

8

2

25

↑



iabvek
@iabvek

...

bet an additional \$15k to [@TarasBob](#)' \$30k



Avraham Eisenberg @avi_eisen · Jan 19

Bet my 10k vs @TarasBob 66.5k that [metaculus.com/questions/8898...](https://metaculus.com/questions/8898/) resolves yes



metaculus.com

Will Russia invade Ukrainian territory before 2023?

3

5

19



Avraham Eisenberg @avi_eisen · Feb 28

Confirming @TarasBob has paid this

2

1

17



Lucas @Talophex · Feb 28

One of the biggest L's I've seen in a while, dude evaporated a whole pile of money.

1

1

1



Taras.eth | Taras.com

@TarasBob

...

Replying to @Talophex and @avi_eisen

Didn't evaporate. The money went to people who were better predictors. 😊



Taras.eth | Taras.com

@TarasBob

...

Just paid [@kiwifruitzzz](#) \$340k. I was confident that Russia would not invade until the last moment. I thought it would be an idiotic move for him to do that. So many risks, and no benefits even if they win. I was betting that Putin was rational. I was wrong.



falseflagfruits 🔒 @kiwifruitzzz · 2/19/22 ...

adding on another 160k vs [@TarasBob](#) 240k for a total of 175k vs 340k. Resolves to yes if either the original criteria resolves yes, or if the similar metaculus question (metaculus.com/questions/8898...) resolves yes.



Will Russia invade Ukrainian territory before 2023?
metaculus.com

1

4

6



falseflagfruits 🔒

@kiwifruitzzz

...

this has been resolved yes and paid by
[@TarasBob](#)

10:48 PM · 2/28/22 · Twitter Web App

TarasBob paid them all. He also happens to have a surprisingly interesting [website \(a\)](#).

Various excellent forecasters wrote a bunch about the Ukraine invasion. Michał Dubrawski collects a bunch of them [here \(a\)](#), but these pieces become outdated pretty quickly. Zvi

likewise covered [various prediction platforms](#) (a) on Ukraine.

The US is facing a helium shortage, and thus [sending fewer atmospheric balloons](#) (a), which could affect weather forecasters. The long-run explanation involves US mismanagement, which led to the US [selling off its reserves starting in the 90s](#) (a). The short-run explanation also involves US mismanagement, but this time also combined with [over-reliance on Qatar and Russia](#) (a).

The [\\$4k Impactful Forecasting Prize](#) is still running until the 11th of March. It has not yet [seen many entries](#), so the expected value of applying seems high.

NVIDIA released some [complex time-series forecasting infrastructure](#) (a) (code [here](#) (a)) to allow testing many different time-series models.

I enjoyed [Ege Erdil's quantified essay on Computability and Complexity](#) (a) on Metaculus.

Long Content

[No One Cared About My Spreadsheets](#) (a). Bryan Caplan, the author of *The Case Against Education*, mentions that nobody criticized the painstakingly-made calculations underlying his book.

[Evident Method](#) (a) was a forecasting training consultancy by the now presumably very busy [Danny Hernandez](#) (a). The website is beautiful, and in a world where I had more time, I might want to take over it.

I'll show that a 20% improvement in identifying upfront which projects are destined to be failures based on cost is tractable (they were going to take so long that the organization would regret starting them if it'd known the true cost).

Note to the future: All links are added automatically to the Internet Archive, using this [tool](#) (a). "(a)" for archived links was inspired by [Milan Griffes](#) (a), [Andrew Zuckerman](#) (a), and [Alexey Guzey](#) (a).

I have their priors, I give them information, I can observe whether they update like a Bayesian would.

— Eva Vivalt

Forecasting Newsletter: April 2022

Highlights

- Keine Davon to become German Chancellor despite prediction markets' confidence to the contrary
- Netflix releases Korean soap opera: Forecasting Love And Weather.
- Hague to allow Treaty on Accuracy to stand

Index

- Highlights
- Prediction Markets & Forecasting Platforms
- In The News
- Long Content
- Hard To Categorize

You can sign up for this newsletter on [Substack](#), or browse past newsletters [here](#). If you have a content suggestion or want to reach out, you can leave a comment or find me on [Twitter](#).

I have received an offer I couldn't refuse from a premier Substack competitor, so this newsletter will be moving to [onlyfans.com/forecasting](#) starting next month (I had some troubles with verification this month). Although I understand that it might be awkward for some readers, the signup bonus alone made this the utility-maximizing move. I am also excited about incorporating OnlyFan's paying functionality to streamline my consulting and allow readers to solicit calibrated forecasts.

Prediction Markets & Forecasting Platforms

Palantir, a controversial (approval rating: 22%, source: Poll aggregation by FiveFourTwo) defence contractor headed by semiquincentennial entrepreneur, past antipope and presidential candidate Peter Thiel, has launched its first assassination market in collaboration with the UN's Security Council. Participants will have the possibility to anonymously bet on the date of the death or disappearance of the elusive globetrotter terrorist and hacker known only as "Morpheus". In an unusually emotional speech, UN Security Council head-honcho Malia Ngo profusely thanked Thiel, saying that it "warms [her] heart to see that human innovation can help contain such disruptions to the normal functioning of civilization."

Ought, the machine learning research lab, has been acquired by Metacortex. Metacortex predicts (confidence: 79%, source: Metacortex proprietary systems) that it will be able to successfully tightly integrate Ought's autonomous research, forecasting and decision-making capabilities into its AI-based defence and deterrence products. Metacortex's stock market valuation rose 0.12% on intra-minute trading after the announcement.

As the Argentina-UCS cold war continues, [Mary Ann Island](#), a small island previously administered by Argentina, has been invaded by a confederacy of independent traders seeking to exploit ambiguity in some prediction markets' resolution criteria. Some high-volume prediction markets were set up to give advance warning of a possible invasion of any part of Argentina but neglected to specify that the invading party had to be the UCS as an exercise in diplomatic tact. The island itself is unpopulated and known for its large population of rabbits, but otherwise unremarkable.

In the News

The International Court of Justice in the Hague has allowed the Treaty on Accuracy, and in particular, its harsh punitive measures, to stand. The Commentators, Litterateurs And Pundits Society (CLAPS) had previously argued that not differentiating between an assertion of fact, an unfounded opinion and a calibrated forecast was a permitted exercise of "free speech", whereas Chief Prosecutor Michael Townsend successfully argued before the court that readers have a symmetric right to true facts and that this right justifies restrictions in journalistic freedoms. To comply with the new regulations, this newsletter shall (probability estimation: 95%, source: personal estimate) here onwards incorporate probabilistic estimates of statements with less than 98% probability; a third party service will ensure and incentivize calibration.

Great Britain's GDP is now 2^{10} times larger than that of continental Europe. Since it replaced its ceremonial monarchy with a futarchy-based decentralized parliamentary system set to optimize "hedons", Great Britain's economy has been doubling every four months, which stands in sharp contrast to an average doubling time of one year in Honduras, one and a half in the Mars colony, two years in continental Europe, five years in developing nations, or ten in the United Catholic States of America. Nonetheless, the methods of The Great DAO of Great Britain remain controversial (50.1% approval rate among eligible voters.) For example, despite Metacortex's highly accurate simulations conclusively (99.9%+) having shown that acting decisively against rebel Scottish separatists was a necessary move to preserve Great Britain's prosperity, a group of revisionist historians recently argued that obliterating Edinburgh with a kinetic orbital strike was "morally wrong" and a display of "excessive force".

Succession troubles in the Arab Emirates intensify, as prediction markets and calibrated proprietary systems predict that a less charismatic brother would reign more effectively than the current heir apparent. Current reigning monarch Abdulaziz bin Salman still holds the power to appoint his heir, but choosing an in-expectation-worse successor might (probability estimate: 75%, source: personal estimation) lead to a loss of legitimacy and public unrest (e.g., protests), but would probably not topple the regime (20% that it will, source: personal estimation.)

As foreseen by prediction markets and pundits alike, Keine Davon has been elected leader of the CDU, and is widely expected to become the German Chancellor in the upcoming elections this June (e.g., FiveFourtyTwo currently gives this a 97% probability). I'd personally give it 95%+ probability, however, prediction markets are currently sitting at 85% because of a small minority of ardently delusional deniers who expect the candidacy to be rendered illegal after judicial review.

UN Secretary-General Yan Zhang vows to move prediction markets to at least a 30% implied probability that the Spanish military junta will not be in power by the end of the decade. Prediction markets rose to 35% upon announcement (source: Metacortex), up from an early estimate of 28%. The move is widely considered to be an attempt by Zhang to distract attention away from an embezzlement scandal, in which famine prediction systems were manipulated to show increasing risk in areas that were actually safe, leading to the deployment of additional funds which could then safely be stolen.



Netflix releases a new Korean soap opera, [Forecasting Love and Weather](#), which tells the gripping tale of how a young man with an affinity and talent for weather forecasting falls in love with an analytical woman of comparable forecasting prowess. "It was as if an occult

hand had reached into Korean society and made forecasting cool and mainstream", mentions a spokesman for the Korean Forecasting Congregation. It further seems that a lot of [attention to detail](#) went into making the show realistic.

Mars Emperor [Tim Chu](#) vows to colonize Andromeda. Prediction markets rose to 99% upon announcement, up from an early estimate of 0.5% (source: Metacortex.)

Recent blog posts

[Sand Teal Cortex](#) investigates the story of the Chinese precogs who are rumoured to have recently been making waves in the prediction and stock markets (quantified in later sentences). In short, in the 2050s, the then-communist Chinese regime started an embryo editing and selection program (99%+; this is well documented) for a variety of traits, i.e., for charisma, military-strategic ability, mathematical talent, etc. Most of these experiments otherwise never went anywhere that we know of (30%, the fact that there isn't public information doesn't update me much either way, and this contradicts [theoretical models](#)). However, after an unknown number of generations, humans optimized for correlates of predictive prowess reportedly displayed truly uncanny predictive ability (70%; reports are unclear, but again theoretical models suggest that gains in the absence of ethical constraints can be massive). After the fall of the Chinese communist regime, these precogs are speculated to have begun to use those abilities for profit (35%; here we enter the realms of speculation). This would—so the theory goes—explain a recent very noticeable upwards blip in the accuracy of various prediction markets.

In particular, since a couple of days ago, global financial markets have begun acting strangely, in a way that suggests that some entity has been exponentially growing the fraction of total market power it controls (40%; I'm deferring to the experts here, but don't have detailed models myself.) Prediction markets on the topic don't have much liquidity yet, but in the meantime, superforecasting systems give [rest of sentence interdicted on the authority of Guardian Samuel Kuehlruhe].

Trigger warning: *Reading the next paragraph is grossly illegal in the UTS and allied jurisdictions. If you're an emulated being, consult your TOS or face termination at your own risk before proceeding. Honestly, I thought that this was worth reporting on, but at least get a VPN, plz.*

Rootclaim has a new feature analyzing the reasons for Peter Thiel's extraordinary longevity. They find that the most likely hypotheses are a combination of cryogenic stasis (75%), speculative medical procedures (85%) (e.g., blood transfusion from younger Thiel clones (45%)), and replacement by clones once the original Thiel becomes too decrepit (35%). One can only hope (20%; informal estimation) that articles such as this will halt—or at least decelerate—the seemly inevitable rise of the Thielian church.

Long Content



[Robin Hanson To Represent Sweden At 2021 Olympic Games In Tokyo](#). To settle a bet about whether he would have found a career in sports more meaningful than his intellectual career, Robin Hanson has agreed to spin up universe afea6ef9628fcb91771abc9f799cf15. You can bet on the outcome [here](#). [United Nations Security Council Resolution 26280](#) requires us to inform you that if there are two or more Robin Hansons in your universe, you might be in a simulation (probability depends on the specific [anthropic question being asked](#) and on how much credence one lends to the [simulation hypothesis](#).)

T. Greer of [The Scholar's Stage](#) speculates (implied probability estimate: 7%, source: Scholar's Stage) that Russia has systematically been misleading US analysts as to the efficacy of various forecasting methodologies. He proposes this as an explanation as to why superforecasters are better at predicting geopolitical events, but monetary prediction markets are better at everything else. The idea is that the KGB would have carried out their own experiments to determine which forecasting method is more accurate, and then changed its own actions in low-stakes events in the geopolitical arena to make superforecasting appear superior, so that its rivals would have access to worse probability elicitation measures in situations where it truly mattered.

This newsletter is generously sponsored by Metacortex and the Cult of Tim Chu. They cover server costs for around twenty subjective hours a month, which is just barely enough to write this newsletter, so I rely on subscribers to exist beyond that. Please become a paying subscriber. Please become a paying subscriber. [Please become a paying subscriber](#).

They said that conquering Afghanistan had been tried before, that it was a fool's errand. But if you are a strong enough optimizer, base-rates don't apply. That's why Afghanistan is now a paradise on Earth, and that's how I got a nation-sized impenetrable fortress.

—Peter Thiel

Forecasting Newsletter: March 2022

Highlights

- [Comparing top forecasters and domain experts](#) finds that past studies mainly were not comparing apples to apples and that the assertion that superforecasters were 30% better than intelligence analysts was unjustified.
- [Samotsvety's Nuclear Forecasts](#) got picked up in the [Spanish press](#) and criticized by a [pessimistic nuclear expert](#).
- [Forecasting Wiki launched](#)
- Polymarket is inflating its volume by incentivizing wash trading. (edit: apparently not the case, will issue a correction in the next issue)

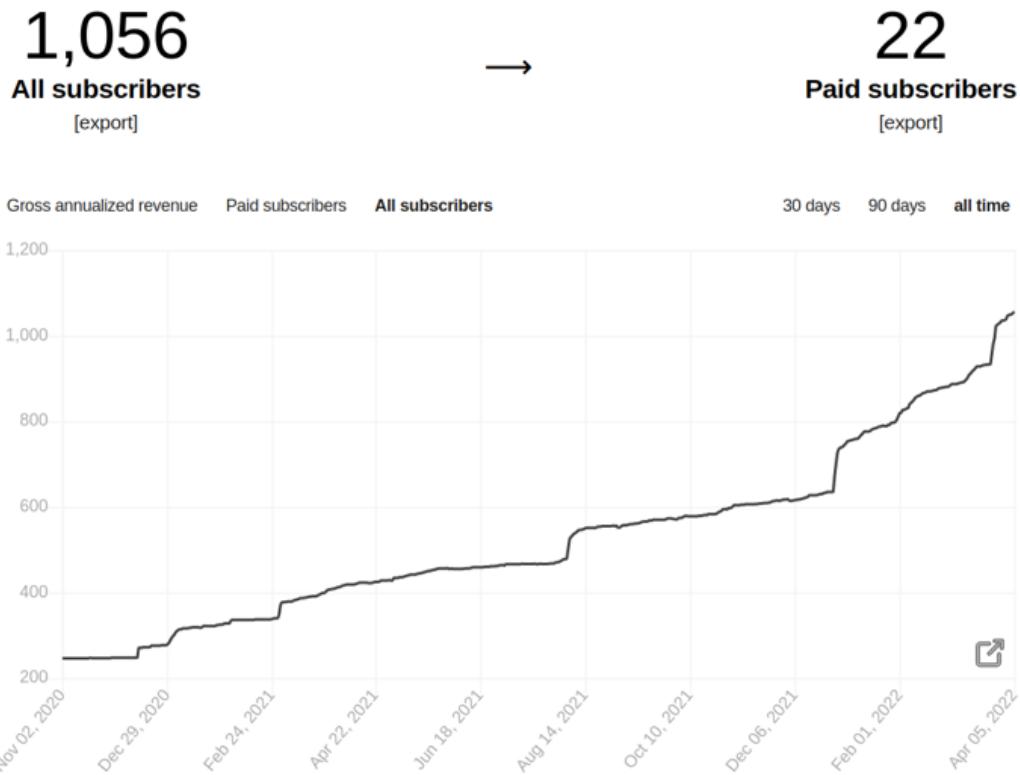
Index

- The state of forecasting
- Notable news
- Platform by platform
- Relevant research

You can sign up for this newsletter on [substack](#), or browse past newsletters [here](#).

The state of forecasting

On account of getting a plug on one of Spain's most-read newspapers, this newsletter has reached 1,000 subscribers:



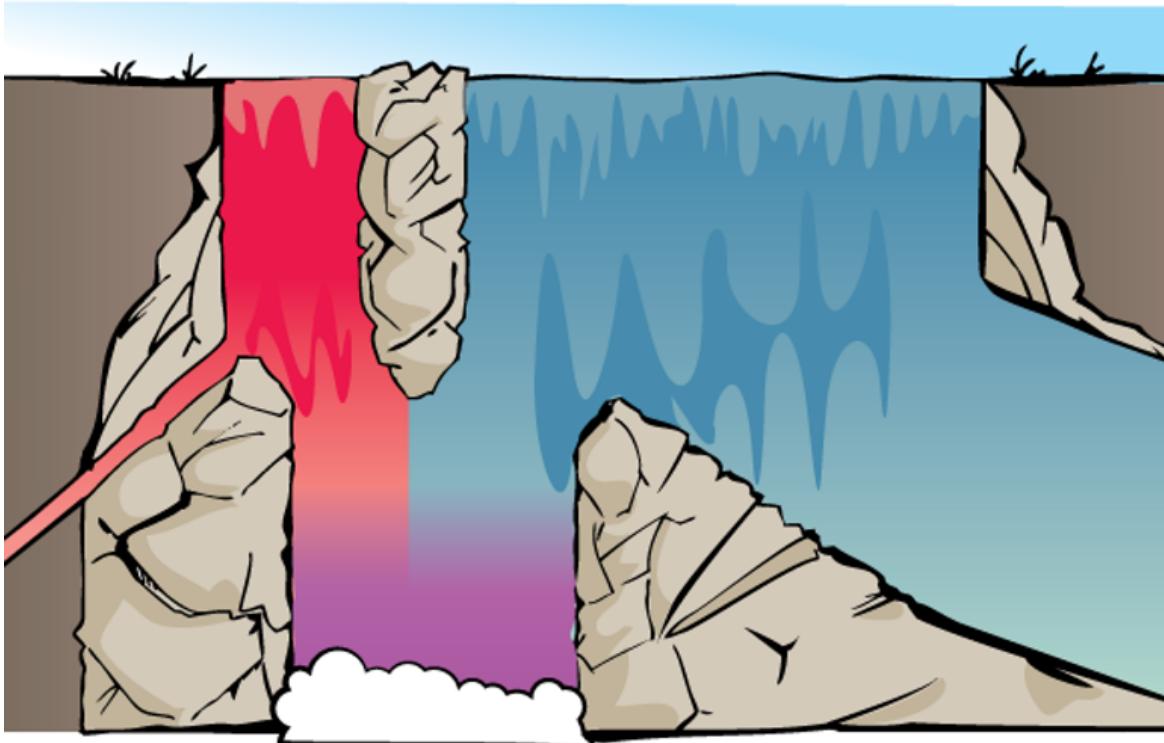
You can find a market on when it will reach 2000 [here](#).

So I thought I would summarize the state of forecasting as I see it, striving to be informative to new readers. If you're already familiar with the key points, you might want to skip to the next section.

The main problem is bullshit or lack of epistemic virtue and ability. The US misled itself into thinking that Iraq still had weapons of mass destruction or that [everything would be okay in Afghanistan](#) (a). People were not expecting covid to last so long. And everyone keeps expecting a better brand of politician to show up.

What is the alternative? The alternative is to develop better models of the world and then use those better models to make better decisions.

But how do we know which models of the world are good? How do we differentiate real understanding from fake understanding? It's tricky, but to a first approximation, we make our hypotheses about the world output predictions, and we [reduce our confidence in the hypotheses that make worse predictions](#) (a). The book *Superforecasting* is a neat introduction to the practices involved. E.T. Jaynes' *Probability Theory: The Logic of Science* is a hardcore introduction to the math behind it. Both books are probably available for free in the [z library](#) (a).



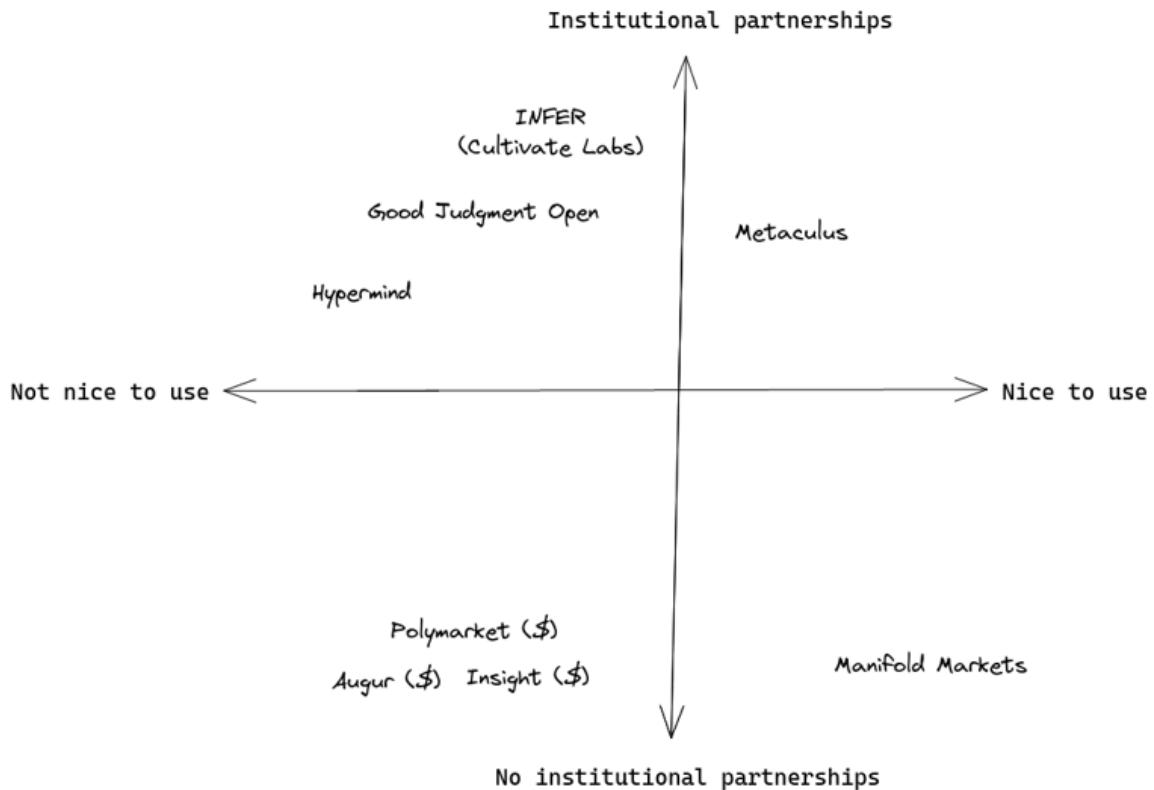
A graphical representation of [Bayes' rule](#), from [Arbital](#).

You could keep track of your probabilities in a spreadsheet. But it would also be convenient to collaborate and compete with others. And here come various forecasting platforms, like [Metaculus](#) (a), [Manifold Markets](#) (a), Good Judgment Open, or INFER. These forecasting platforms struggle to seduce forecasters into tracking their probabilities on their site and get the funds of decision-makers who want to use probabilities to make better decisions.

Besides forecasting platforms, we also have real-money prediction markets, where participants bet their own money on their degree of belief. These can either be based on cryptocurrencies, like [Polymarket](#) (a), [Insight prediction](#) (a), [Hedgehog](#) (a), or be regulated, like Betfair, Kalshi, Nadex or PredictIt. Historically, prediction markets have focused on sports, but in recent times, they have also hosted more informative markets, e.g., on covid, the invasion of Ukraine, and various US political developments.

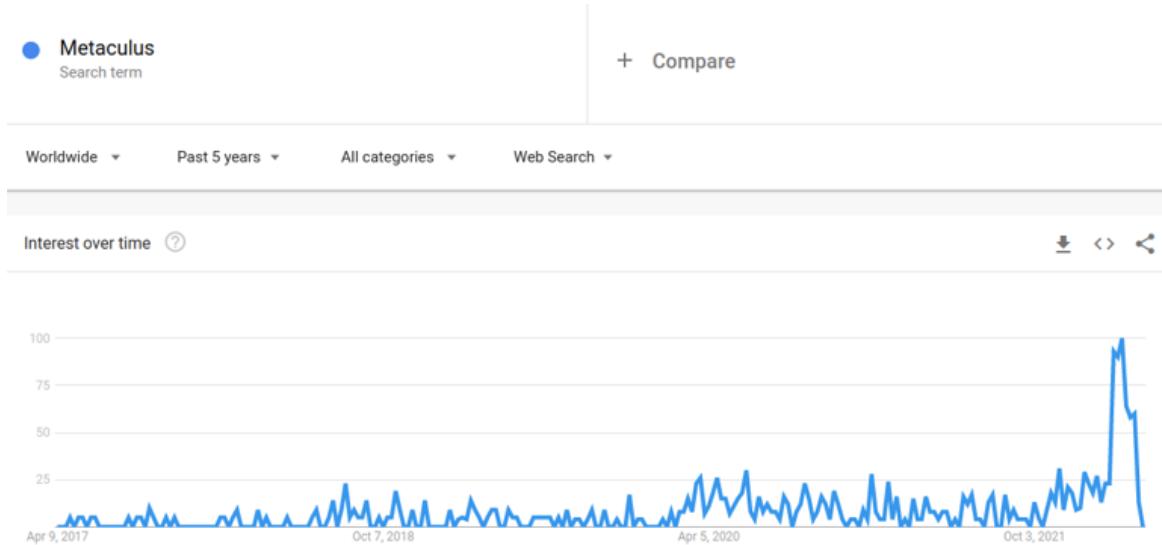
To my new Spanish readers, I would recommend that you start forecasting on [Metaculus](#) and only consider trying prediction markets once you've proven to be good in platforms that don't risk real money.

Something that has been on my mind is that forecasting platforms tend to either have institutional partnerships or be nice to use. But generally not both. I think this can be explained by older websites using worse technology but having had more time to develop partnerships:



I generally tend to take a *technology maximalist* perspective toward that tradeoff in this newsletter. I tend to express the view that platforms with better technology will outcompete the others because they will be able to move and experiment faster, add new features, and retain more users.

Recently, two interesting developments have been affecting the forecasting ecosystem. First, the war between Russia and Ukraine has sparked broader interest in whether forecasting platforms or prediction markets have anything to say about it:



Popularity of the search term "Metaculus" in Google trends. h/t Metaculus user [UgandaMaximum](#)

And secondly, the [FTX Future Fund](#) (a), a very large philanthropic funder, has expressed interest in forecasting. Platforms and individuals in the space have been scrambling to present proposals that might please it.

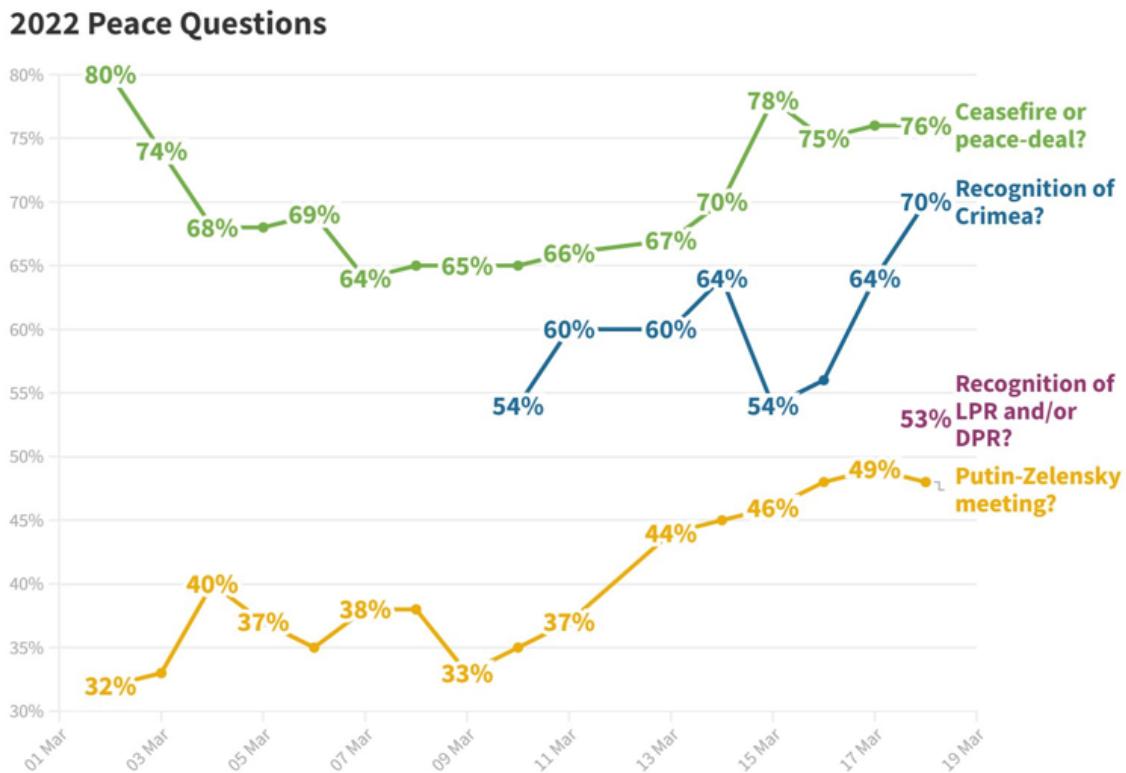
And with this, we are left to discuss recent developments:

Notable news

[Pricing existential risk](#) (see also: [existential risk](#) (a)): All investments go to zero in the case of existential risk, so it's hard to price it correctly. In particular, one can't just substitute riskier assets with less risky assets. Still, the higher the existential risk is, the more one should frontload consumption. And if stocks are roughly worth the discounted value of dividends and other payments, higher existential risk should reduce their value. But the market may not have realized this yet. I thought that the article was great, but I would have appreciated a more comprehensive treatment.

[The Forecasting Wiki](#) (a) is getting started. As advertised on their website, they have a meetup on April 24th, as well as a Discord channel.

[Global Guessing](#) continues to do a great job following developments in the Ukraine war through shifts in probabilities. For example:



Global Guessing's tracking of probabilities about the Ukraine conflict.

Platform by platform

Metaculus [continued publishing questions on the Ukraine conflict](#) (a), [estimated low meat production](#) (a) and organized a [small White Hat cybersecurity tournament](#) (a), which got picked up by [Lawfare](#) (a)

Per SimonM, the most insightful comments on Metaculus were:

- [orion.tjungarryi](#) looks at the relationship between population and how long cities hold out, to figure out whether Kiev would fall. The larger the cities, the longer they tend to hold.
- [haukurth](#): "It's a full time job now to constantly degrade Russian chances on various Metaculus questions."
- [aqsalose](#) calculates a base rate for regime change in Russia. Based on historical precedent, Putin's grip on power doesn't look to bad in the short term.
- [Joker](#) also looks at the base rate of sieges—they last longer than a month. Based on this, he gave a 1% chance of Kiev falling at a time when the Metaculus aggregate was at ~65%.

I also liked Richard Hanania's Metaculus notebook on [Why Forecasting War is Hard](#) (a).

Good Judgement Inc is hiring a [Director of Sales](#) (a).

Manifold Markets [discusses their market mechanics](#) (a) (technical). Prediction markets need a way to match bets between users. In modern times, they do so by betting against a central automated market-maker, but different algorithms determine the specifics. Manifold Markets tells how they started with Dynamic Parmimutuel, considered the logarithmic market scoring rule, and ended up with a less elegant constant product market maker.

Manifold also [implemented loans on the first M\\$20 bet on any market](#) (a), [applied to the FTX Fund](#) (a), and [awarded some bounties to active community members](#) (a).

INFER is organizing a tournament for [EA university groups](#) (a). I would recommend joining; I enjoyed their team functionality.

[Insight predictions](#) (a) continues to have the guts to ask the important questions, such as: "Will Russia Conquer the Donbass by the End of July 2022?". Though liquidity (the opportunity to trade on both sides of a question) is a bit thin.

The screenshot shows a prediction market interface for the question "Will Russia Conquer the Donbass by the End of July 2022?".

Contract: Donbass Conquered

Latest Yes Price: 61¢

Options: Yes (green button) and No (grey button)

Bid/Ask Table:

Sell Yes		Buy Yes	
Bid price	Bid size	Ask price	Ask size
56	100	60	100
55	400	61	200
54	200	64	100
50	200	65	243
		70	400

Buy/Sell Buttons: Buy (blue) and Sell (grey)

Input Fields: Maximum Buy Price (60), Number Of Shares (100)

Note: The currency of this market is USD. 100 cents = 1 USD.

Buttons: Next (blue), Skip Offer Confirmation (checkbox)

The ¿founder? of Insight Predictions also [objected](#) (a) to me characterizing Insight as possibly but most likely not a scam in a previous newsletter. One of the key elements that made me suspicious was that he had previously remained anonymous. But he has now de-anonymized himself, and he turns out to be [Douglas Campbell](#), who previously served in Obama's Council of Economic Advisors. So there's that.

[Kalshi](#) (a) and [Polymarket](#) (a) offer markets on interest rate hikes by the US Federal Reserve. This seems like an interesting hedge.

Hypermind has a small [\\$5k tournament on African developments](#) (a)

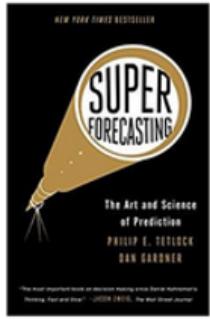
A screenshot of a Twitter post from the account @PolymarketHQ. The post features a blue circular profile picture with a white geometric logo. The text reads: "6,900,000,000" followed by "6.9 **B**illion outcome shares have been traded on Polymarket." Below this, the text "Our sexiest accomplishment to date." is followed by "To celebrate, we're doing a lil giveaway..." with a yellow hand icon pointing down. At the bottom, it shows the timestamp "2:49 PM · Mar 30, 2022 · Twitter Web App" and engagement metrics "119 Retweets 5 Quote Tweets 175 Likes".

Polymarket has been offering rewards for trading. Trading incurs a fee, but trading rewards are higher, which incentivizes wash trading (trading back-and-forth at high volumes.) The thing is, Polymarket developers are not stupid, so I'm guessing that they are doing this because they want the volume to be as high as possible ¿possibly to impress or appease investors? The non-nefarious explanation is that they deeply want to attract new traders and keep the engagement of old ones, and are ok paying wash traders as the cost of doing business. (edit: apparently not the case, will issue a correction in the next issue)

In any case, I have downgraded [my estimates](#) (a) of Polymarket prediction quality as a function of volume for Metaforecast. [Metaforecast](#) (a) itself is doing great, with a bit over 15k views a month. I've also recently hired an [extremely competent developer](#) (a) to continue working on the project. So far, he has been leaving the codebase in a much better position, solidifying and professionalizing parts that were previously more glued together with ducttape. [Feature ideas](#) are welcome!

[Spose](#) (a) (pronounced like "I suppose", I'm guessing) is a smallish platform to "casually forecast serious stuff". They ask one very short-term question every day.

Research



Who are the Superforecasters?

Good Judgment's global network of Superforecasters has its roots in research funded by the US intelligence community. Reports that Superforecasters were 30% more accurate than intelligence analysts with access to classified information rocked the conventional wisdom.

Source: goodjudgment.com frontpage.

[Comparing top forecasters and domain experts](#) (a) reviews the idea that the very best generalist forecasters can beat experts at predicting events *in their own domain of expertise*.

In particular, there is an oft-cited refrain that "superforecasters are 30% better than experts with access to classified information". But the authors find that a large share of the difference may boil down to different aggregation methods: "*The forecaster prediction market performed about as well as the intelligence analyst prediction market; and in general, prediction pools outperform prediction markets in the current market regime (e.g. low subsidies, low volume, perverse incentives, narrow demographics).*"

The CEO of Good Judgment Inc answers [in the comments](#) (a): "*These claims about Superforecasting are eye-catching. However, it's difficult to draw any conclusions when most of the research cited doesn't in fact include Superforecasters*". But this seems inconsistent with the eye-catching 30% claim on Good Judgment's own website.

My forecasting group recently estimated the [risks of nuclear war](#) (a). We arrived at a 24 in a million chance that an "informed and unbiased" Londoner would be hit by a nuclear blast in the next month. This estimate was picked up by [Scott Alexander](#) (a) and the [Spanish press](#) (a)

Now a subject matter expert who served as deputy staff director of the Senate Committee on Foreign Relations where he worked on approval of the New START agreement, [criticized our estimates](#) (a). Our answer can be seen in [the comments](#) (a).



Peter Wildeford (hiring 0/25!) @peterwildeford

Mar 26

Help the respected nuclear domain experts and the respected generalist forecasting experts are fighting again



Dr. Jeffrey Lewis ✅ @ArmsControlWonk

Mar 26

Contra a recent, weird “forecasting” assessment about the risks of nuclear use against London arising from the war in Ukraine, this piece by @PeterScoblic is excellent.

[forum.effectivealtruism.org/...](https://forum.effectivealtruism.org/)

Mar 26, 2022 · 7:10 PM UTC

1 reply 0 retweets 9 likes

[Why short-range forecasting can be useful for longtermism \(a\)](#)

I argue that advances in short-range forecasting (particularly in quality of predictions, number of hours, and the quality and decision-relevance of questions) can be robustly and significantly useful for existential risk reduction, even without directly improving our ability to forecast long-range outcomes, and without large step-change improvements to our current approaches to forecasting itself (as opposed to our pipelines for and ways of organizing forecasting efforts).

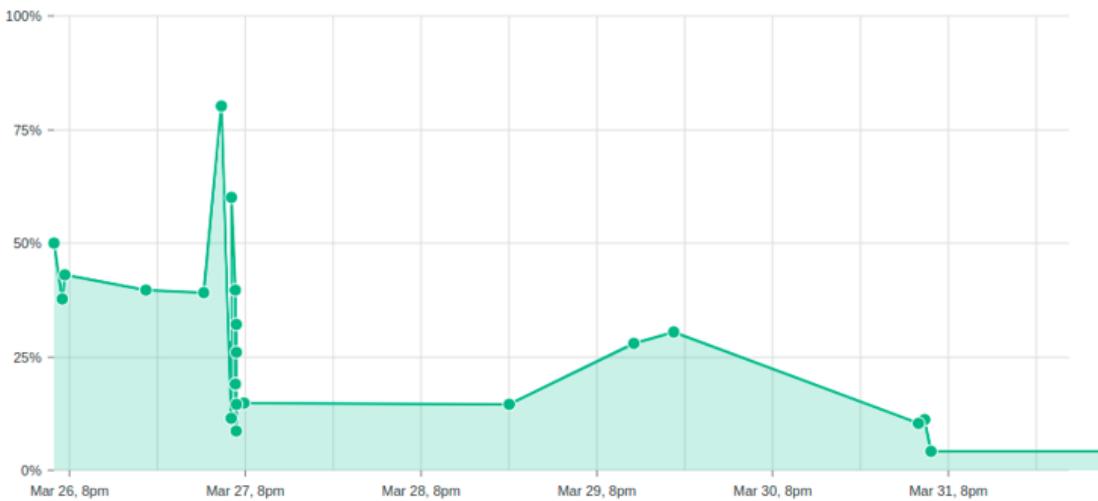
To do this, I propose the hypothetical example of a futuristic EA Early Warning Forecasting Center. The main intent is that, in the lead up to or early stages of potential major crises (particularly in bio and AI), EAs can potentially (a) have several weeks of lead time to divert our efforts to respond rapidly to such crises and (b) target those efforts effectively.

In [Cryptoepistemology \(a\)](#), davidad maps different theories of justified beliefs to different styles of cryptographic proof.

Conditional on the next character in the following post being "§", will the post be curated on LessWrong?

4%
chance

Matthew Barnett Mar 26 - Apr 11 (in 9 days) EDIT M\$ 1,564 volume Tweet Embed



Tags #LessWrong #personal

Type a tag...

SAVE TAGS



Matthew Barnett created this market 6 days ago

Post body: "This post is not what you think it is. It is a threat. If LessWrong moderators do not curate this post, I will take immediate action against them, includ"

ADD TO DESCRIPTION



Lastly, I really enjoyed two prediction-market related April Fool's jokes: [Using prediction markets to generate LessWrong posts](#) (a) and [Anti-Corruption Market](#) (a). I'm also pretty proud of my own April Fool's: [Forecasting Newsletter: April 2222](#) (a).

Note to the future: All links are added automatically to the Internet Archive, using this [tool](#) (a). "(a)" for archived links was inspired by [Milan Griffes](#) (a), [Andrew Zuckerman](#) (a), and [Alexey Guzey](#) (a).

y en el mundo, en conclusión,
todos sueñan lo que son,
aunque ninguno lo entiende.

English translation:

and in the world, in conclusion,
they all dream what they are
although none of them understands it

Fragment of Segismundo's monologue, in *La vida es sueño*, from Spanish playwright Calderón de la Barca.