

Forecasting 101

Jacob Gloudemans

Forecasting 101

Copyright © 2025 by Jacob Gloudemans

All rights reserved.

No portion of this book may be reproduced in any form
without written permission from the publisher or author,
except as permitted by U.S. copyright law.

Contents

Introduction	4
I. Foundations	6
What is a forecast?	7
Where do forecasts come from?	9
What makes a forecaster good?	10
Crowd forecasting	15
State of play	21
What is forecasting good for?	23
Your first forecast	28
II. Strategies and Skills	33
Practice	34
Measuring success	36
Writing	44
Base rates	50
Minimum viable models	53
Intuitive forecasting	57
Doing your own research	61
Updating	65
Cognitive biases	67
How to not lose all your money	75
Recommendations for Further Reading	90

Introduction

In July 2023, I joined a ‘forecasting tournament’ hosted by the website *Metaculus*. I had been vaguely familiar with prediction markets for years and had recently finished Philip Tetlock’s book *Superforecasting* which documents how he and his team identified, and later trained groups of elite forecasters. I was between jobs and had too much time on my hands, and it seemed like a fun thing to try. So I signed up for Metaculus and joined their Quarterly Cup, a 3 month tournament in which participants predict how breaking news events will unfold.

I remember being nervous to click submit on my first few forecasts. This seems laughable now, with thousands of predictions to my name, including some with real money on the line. But at the time I had almost never posted anything on the internet, so perhaps I was just nervous to put my opinions out into the world, even in a relatively obscure little corner.

I won the tournament. And then I entered the next Quarterly Cup and won again. Then I entered a much larger tournament, with thousands of participants and with money at stake, and finished 3rd. I ended 2023 ranked #3 across all forecasters on Metaculus. On another prediction platform, Manifold Markets, I grew my initial 500 ‘Mana’ to over 50,000 in 6 months.

I wasn’t sure what to make of this success. I’d worked hard for these results (in the first tournament I wrote about almost every question, publicly on the internet), and took detailed notes on each success and failure. Whenever I made a forecast I would update it religiously, sometimes multiple times per day. But I also had no prior experience. I was making predictions on topics I often knew nothing about.

One of the key lessons of *Superforecasting* is that good judgement is a learnable skill. I agree with this. I think most of my success came from simply doing the work, researching every question, admitting and growing from my mistakes, and working hard to counteract my biases. These are practices anyone can emulate.

At the same time, it's not a skill that you learn once and keep forever. It requires consistent effort and continual learning. The hardest part isn't 'knowing the strategies to make a good forecast,' it's applying those strategies every single time, even when you're exhausted, or don't want to do the research, or when placing a risky, impulsive bet feels more fun. It's no coincidence that my worst run of forecasting came when I was busy with a full time job, stressed about life, and forecasting an event where I had a deep emotional attachment to the outcome.

In this guide, I explain what has made me a successful forecaster, and offer practical advice to others who find themselves where I was two years ago, trying their hand at forecasting and striving to make better predictions about the future. I assume minimal prior knowledge, beginning with a review of the foundational ideas behind crowd forecasting, before detailing the strategies and skills that I've used to become a better forecaster.

I. Foundations

A primer on crowd forecasting

What is a forecast?

A *forecast* is a probabilistic prediction about the future. Everyone is familiar with the idea of a weather forecast. “There’s a 70% chance of rain on Tuesday” is a probabilistic prediction about the chances of rain on Tuesday. We can make a forecast about any type of future event:

- The Packers have an 80% chance of beating the Bengals this weekend
- Donald Trump has a 65% chance of winning the election
- There’s a 7% chance China will invade Taiwan in the next 5 years
- There’s a 90% chance the Federal Reserve will lower interest rates at the next meeting
- There’s a 22% chance OpenAI will release GPT-6 in 2026

Forecasts are useful because the future is *uncertain*. All manner of things *could* happen in the world every day, every week, and every year, things that would matter in our lives should they happen. There could be an economic downturn. Housing prices could rise sharply, decline, or stay flat. The country could go to war. You could win the lottery.

Forecasting helps us quantify our uncertainty about the future, which in turn allows us to make the best decisions possible, given the incomplete information we have. They reveal both the *absolute* chances of things (should I be worried a lot, a little, or not at all about that thing?), as well as the *relative* chances of different possibilities (which outcomes are more likely and how should I allocate my resources?). Here are a few concrete examples:

- You want to go for a hike this weekend. You can go on either Saturday or Sunday and you'd like to pick the day that's most likely to have nice weather. By observing a weather forecast, you can increase the chance that you'll have a pleasant hike.
- You're a government employee, or work at a company that depends on government employees. A government shutdown would directly impact you or your business, and you've heard in the news there's a possibility of one happening soon. Forecasts about the likelihood (and duration) of a shutdown can help you understand how seriously you should be preparing for one.
- You're a voter who's concerned about the Russia-Ukraine war or another geopolitical conflict. Forecasts about the likely foreign policy outcomes under each candidate's leadership help you vote for the candidate most likely to achieve your preferred outcome.
- You're a philanthropist concerned with existential risks to humanity. You've heard that pandemics, meteor strikes, climate change, and runaway artificial intelligence all pose risks to humanity. By forecasting the likelihood of human extinction from each of these causes (absent further intervention), you can understand the relative severity of each risk and direct your donations towards the most significant one.

Where do forecasts come from?

Forecasts can be based on quantitative analysis, relying on numerical data and statistical models, or can lean on non-numerical judgements and intuition. Many forecasts will blend these two approaches.

Which approach is used depends on the topic and the skills of the forecaster. Domains with large amounts of data, frequent repeated trials, and that depend on well-studied physical processes are especially conducive to quantitative analysis. Weather forecasting is entirely quantitative, using complex physical models running on supercomputers to simulate atmospheric conditions days into the future. Likewise, election forecasting models typically blend a variety of polling data, economic indicators, historical results, and other measurable factors to determine each candidate's chance of victory.

Questions that depend on individual human decisions or are poor in data require a more qualitative, intuition-based forecasting approach. Geopolitical forecasting and domestic politics ("Will country X invade country Y?", "Will Trump attempt to run for a third term?", "Will there be a government shutdown?") are like this.

Often, forecasts will combine quantitative and qualitative approaches. To make predictions about how AI capabilities will progress, we can combine data trends from recent years (performance benchmarks, data center construction, money invested) with technical knowledge and intuition about what bottlenecks or constraints the technology is likely to encounter in the coming years.

What makes a forecaster good?

The best forecasters are both *trustworthy* and *insightful*.

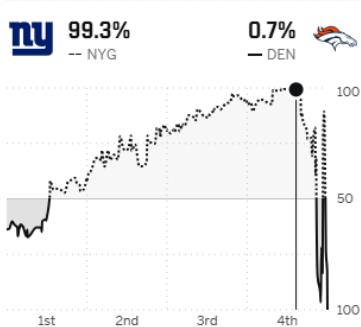
Trustworthiness (often called *calibration* or *reliability*) means that when the forecaster says there's an X% chance of something occurring, that thing really does happen X% of the time. A weather forecaster is trustworthy (or 'well-calibrated') if when you look back at the last 100 times they said there was a 20% chance of rain, it actually rained about 20 out of those 100 times.

This can be counterintuitive. If a forecast says there's only a 10% chance of rain and then it rains, we're tempted to conclude that the forecast was terrible. But with probabilistic forecasts, we can only properly evaluate forecast quality after many trials. If it *never* rains when the weather forecast gives a 10% chance, the forecaster is not actually trustworthy. And if we can't trust the forecast, we can't rely on it to help us make decisions, such as whether to go for a hike or not.

Consider the roll of a fair 6-sided die. We know there's a 16.67% (%) chance of each number being rolled. If we roll the die and it comes up '1', we don't assume the die is weighted and that the 16.67% 'forecast' was wrong. Rather, we understand that we need to roll the die many times to get a sense for the probability of each outcome. In this case, if we roll 100 or 1000 times, we'll see that in the long run, each side does indeed come up around 16.67% of the time.

The other important characteristic of a forecaster is what I call *insightfulness*. Insightful forecasters are *consistently more confident in the eventual outcome, when compared to other forecasters or to simple baselines*. Given two well-calibrated forecasters, we prefer the one who has a track record of insight, the one who makes predictions closer to 0% for events that don't occur, or closer to

WIN PROBABILITY



With 6 minutes left in this NFL game, ESPN Analytics gave the Giants a 99.3% chance of victory, but the Broncos pulled off a stunning comeback to win the game. Was the forecast overconfident? Perhaps, but there are 272 games in an NFL season. If these forecasts are trustworthy, we should expect a team to squander a 99% win probability a few times each season. With 121 games played as of this writing, 2 have seen a team with a 99% win probability go on to lose the game, not far out of line from what we'd expect.

100% for those that do. Sometimes, this means giving a confident prediction when others are wishy-washy, a 90% chance when everyone else says 60%. Other times, it's the opposite, admitting it's a 50/50 tossup when everyone else is confident in the outcome.

Let's revisit the weather example. We can get a well-calibrated forecast for the chance of rain on any given day by looking at the historical rate of rain on that day over many prior years, the 'climatology.' But clearly we can do much better than this. The climatology forecast doesn't look at current atmospheric conditions, which give us useful information about the weather we should expect over the next several days. If there are no clouds in the sky and no fronts moving through, there's a low chance of rain, even if it has rained frequently on this day historically. The forecast on your phone's weather app (based on sophisticated real-time weather models) and the climatology forecast are both trustworthy over the long run, but the weather app is more insightful - it's consistently more confident in the eventual outcome than the climatology forecast.

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Climatology	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%	30%
Weather App	80%	20%	20%	20%	20%	90%	10%	20%	10%	10%	20%	10%	70%	50%
Rain?	Y	N	N	N	Y	Y	N	N	N	N	N	N	Y	N

Two weather forecasts, each giving the probability of rain each day for two weeks. Both forecasts are trustworthy - in the long run, it actually rains roughly as often as each forecast predicts. But the weather app forecast is more insightful, predicting the true outcome with higher confidence on most days.

Unlike trustworthiness, insightfulness (sometimes called *skill*) is a *relative* measure. Climatology weather forecasts are more insightful than random guessing, while modern weather model forecasts are more insightful than climatology.

Using a relative measure is important because some events are more difficult to predict confidently than others. If you can correctly predict the winner of an NFL game 55% of the time, that's unimpressive - a simple rule like 'pick the team with the better record' will do this. On the other hand, if you can predict coin flips 55% of the time, you've got the best forecast on the planet. To evaluate a forecast, we need to know how it stacks up against any obvious baselines, publicly available alternatives, and 'conventional wisdom.'

As a rule of thumb, the best forecasters are *as insightful as possible, while remaining trustworthy*. If a forecaster isn't trustworthy, we have no reason to believe them when they stray from conventional wisdom. Among trustworthy forecasters, the ones who correctly deviate from conventional wisdom most often provide us with the greatest value.

Notice that both trustworthiness and insightfulness can only be established after a forecaster has made many forecasts, and the true outcomes of the events in question are known. Such is life

when we deal in probabilities. Just as you can't judge whether a coin is fair by flipping it once, a single prediction reveals little about the skill of a forecaster.

Trials	% Times # was Rolled						Forecast Quality		
	1	2	3	4	5	6	100% Chance '2'	16.7% Chance Each	
1	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	Great	Bad	
10	30.0%	40.0%	10.0%	20.0%	0.0%	0.0%	Okay	Okay	
100	23.0%	15.0%	16.0%	13.0%	18.0%	15.0%	Bad	Good	
1,000	18.5%	16.4%	17.8%	16.2%	15.9%	15.2%	Bad	Great	
10,000	16.7%	16.4%	16.6%	16.9%	17.0%	16.4%	Bad	Great	

Evaluating forecast calibration and skill requires many trials. A forecast like 'always predict a 100% chance of 2' can look great after one lucky roll, but repeated rolls show that this model is poorly calibrated. After a few dozen trials, it becomes clear that the 'Predict a 16.7% chance for each roll' forecast is the better one.

In 2016, election analyst Nate Silver's¹ final pre-election forecast gave Donald Trump a 30% chance of defeating Hillary Clinton to win the US presidential election. Was this a good forecast?

At a glance, you might say 'No, of course not! Trump won the election!' But I would argue this was actually a *great* forecast. To understand why, let's consider the two measures we've discussed, trustworthiness and insightfulness.

First, Nate has forecasted a large number of elections over the years, and if we look at his results over the long run, we find that his forecasts are well-calibrated. When he says there's a 30% chance of a candidate winning, we can trust that across many elections, such candidates really will win around 30% of the time. Some pundits did predict a Trump victory (although frankly not *that* many), with varying degrees of confidence, but none of them had the long-term track record that Nate did. Though ultimately

¹ Founder of *FiveThirtyEight*, now publishing his work at *Silver Bulletin*

proven correct, their forecasts were not especially useful because we had no idea ahead of time if we could trust them or not.

Second, Nate's 30% forecast was insightful. It still favored Clinton, but much less so than competing election models, betting markets, or the 'conventional wisdom' of the news media, all of which were more confident in a Clinton victory. The New York Times gave Trump a 15% chance of winning, while the Huffington Post gave him a 2% chance. Not only can we trust Nate's election forecasts in general - he's also shown that when he deviates from the crowd, we should listen.

Crowd forecasting

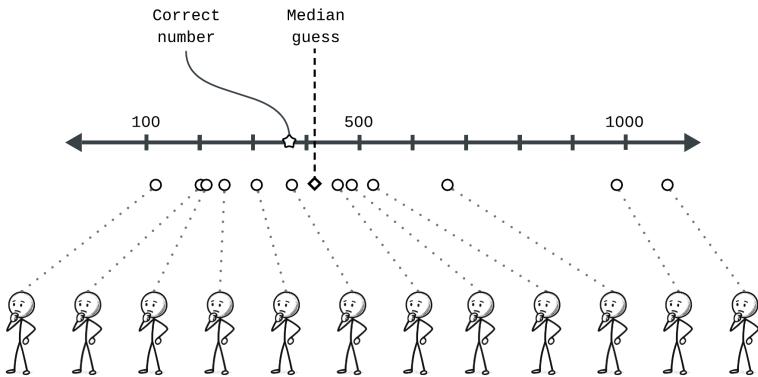
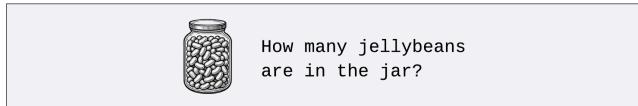
While computation-intensive weather prediction and bespoke election models are indispensable, the space of questions we'd like to apply forecasting to is vast, too vast for us to expect a sophisticated model to be built for each one. And as I mentioned earlier, some domains lack the type of numerical data and historical record that drives such models. Over the last few decades, and particularly over the last 5 years, *crowd forecasting* has emerged as an approach to produce high quality forecasts for virtually any question.

Crowd forecasting is based on the idea of ‘wisdom of the crowds,’ the notion that by combining predictions from many individual forecasters, we can yield a single aggregate forecast that outperforms most or all of the individual forecasts. Each individual forecaster brings a mixture of insight and error. The insights converge on a reasonable guess, while the errors average out to zero.

There are two main approaches to crowd forecasting in widespread use today: *prediction polls* and *prediction markets*.

Prediction polls

In a prediction poll, each participant gives their best guess for the probability of a particular outcome. The individual guesses are then combined into a single ‘community forecast’. The community forecast can be the median of the individual predictions, or it can come from a more sophisticated algorithm, perhaps giving extra weight to predictions by forecasters who were more accurate on previous questions. Forecasters can generally *update* their



The ‘wisdom of the crowds’ at work ... If we collect lots of independent guesses for the number of jellybeans in the jar and take the median, we end up with an estimate that’s close to the correct number, closer than most of the individual guesses.

predictions whenever they choose, with the community forecast giving more weight to the most recent forecasts.

On platforms that implement prediction polling, users ‘win’ by being consistently more confident in the correct outcomes across many questions than the broader community. The highest performing users gain reputation (in the form of rankings or awards) and in some cases can earn money by winning tournaments. Scoring systems use *proper scoring rules* which ensure that users maximize their scores by predicting their true beliefs - they can’t game the system by making lots of bold predictions, because overconfidence is heavily penalized. These scoring systems encode the characteristics of good forecasts we covered earlier. You earn more points by being insightful; if your forecasts

aren't perfectly calibrated, you can always improve your score by improving your calibration.

Prediction markets

In a prediction market, users buy and sell shares of Yes / No outcomes and redeem each share for \$1 if that outcome occurs (or \$0 if it doesn't). The current 'share price' for a given outcome (which ranges from \$0.00 to \$1.00) reflects the crowd's belief about the likelihood of that outcome. Individual users buy or sell shares when they believe the current price for an outcome differs from its true probability of occurring.

For example, consider the question "will OpenAI release GPT-6 in 2026?" If the market-implied probability is 33% (Yes shares are trading at \$0.33, No shares are trading at \$0.67), but you believe the actual probability is more like 15%, you can earn money (and 'correct' the odds) by purchasing No shares. If your forecast is accurate, you're buying shares that are worth \$0.85 for a price of \$0.67.

Thus, prediction markets arrive at reasonable forecasts by providing monetary rewards to users who can consistently make winning trades, i.e. users who improve the prediction market 'forecast.'

Tradeoffs between polls and markets

Prediction markets and prediction polls both can produce trustworthy, insightful forecasts at scale across a broad range of topics. Each approach has its own good and bad characteristics.

Simplicity of use

In prediction polls, users always input their true probability estimate for each question. This simplifies forecasting for the user - all they have to do is come up with a forecast for each question and submit it directly. If their forecast changes, they can update their prediction to the new number.

This has the added bonus that it allows these sites to show the variance in the individual forecasts that are contributing to the community forecast. This makes it possible to distinguish between a 50% forecast where every individual guess is very close to 50%, and a 50% forecast where forecasts range all the way from 10% to 90%.

In a prediction market, traders don't specify their true belief, but rather provide a signal of 'too low' or 'too high' to the market by buying or selling shares. Forecasters must determine how much to invest in any one question based on their confidence, risk appetite, and the size of their bankroll. Compared to prediction polling sites, it's more complicated for forecasters to properly 'express their belief' on a prediction market. As I'll cover in more detail later, it's possible to correctly judge when markets are under or over-priced, but still lose money over time by placing bets that are too large.

Speed

Prediction markets react very quickly to new information. When new information emerges that changes the likelihood of an outcome, there is a strong and immediate financial incentive for traders to incorporate that new information and correct the market

price. In practice, it's extremely difficult to 'win on speed' when trading in prediction markets because this incentive is so strong.

Prediction polls can be comparatively slow to update. Each user's latest prediction is always incorporated into the community average, so even if some users change their forecast in response to new information, a small number of participants who are slow to update can slow down the rate that the community average updates at. Aggregation algorithms can and do give extra weight to recent forecasts, but if this is done too aggressively, the most active forecasters will drown out everyone else.

Information sharing

In a prediction market, once a user has purchased the shares they wish to purchase at a given price, there's no reason (in most cases) for them to withhold information from other traders in the market. In fact, it may be in their interest to share that information, so that the price moves in their favor, and they can sell their shares at the new price for a profit.

In a prediction poll, there is no such incentive, and in fact there is generally an incentive to *withhold* useful information from the community. Scoring generally rewards forecasters for being more correct than the field, *over an extended period of time*. If you believe you have an edge on a question because you've noticed something the community hasn't, your incentive is to hold on to that information for as long as possible. Arguably, there's even an incentive to *actively mislead* the community.

Interestingly, these theoretical incentives do not reflect the commenting behavior you'll find in the real world. The most prominent prediction polling platform, Metaculus, has by far the most insightful, thoughtful comments of any of the major crowd forecasting platforms. The comment sections on real money

prediction markets like Polymarket and Kalshi are generally indistinguishable from what you'd find in the darkest corners of your email spam folder.

I suspect that this reflects the types of users that each platform attracts. There's minimal opportunity to earn money on Metaculus, so the users are there 'for the love of the game' and tend to be the types of people who enjoy writing long, thoughtful comments about obscure topics, even if it technically reduces their edge in a low-stakes tournament.

Rewards

Prediction markets reward good forecasting by redistributing money to the participants who consistently outperform the market, and thus improve the market 'forecast'. Prediction markets are *zero sum* - every penny gained by one trader is directly lost by another trader. Thus, in addition to rewarding skilled forecasters, prediction markets harshly penalize bad forecasters.²

Prediction polling platforms reward users with prestige and credentials ('Superforecaster' or 'Pro Forecaster' are two such accolades), as well as through prize pools for forecasting tournaments. These prize pools tend to be small (relative to the money available on prediction markets) and the expected payoff, even for the best forecasters, is low compared to the time commitment required. This is one reason why prediction polling platforms operate at a much smaller scale than the largest prediction markets today.

² Over time, we'd expect the most skilled traders to dominate prediction markets, as money is redistributed from the worst forecasters to the best ones. This raises an interesting question... if everyone left participating in the market is skilled, can anyone make money? If markets are fairly priced, it should be impossible for anyone to consistently make winning trades. The more accurate a prediction market gets, the less attractive it is for anyone to trade in that market

State of play

While crowd forecasting platforms have existed for several decades, the last few years have seen massive growth in users, influence in public discourse, and investment. Here, I'll survey the most prominent forecasting platforms (as of this writing in fall 2025) and briefly describe what sets each one apart:

Kalshi - a pure prediction market operating legally in the United States (and elsewhere in the world). Kalshi allows users to buy and sell ‘event contracts’ covering a range of topics including politics, economics, technology, sports, and pop culture. It earns money through transaction fees imposed on each trade.

Polymarket - another prediction market, cryptocurrency-based and not currently available in the United States (though as of this writing, it is expected to become available there imminently). Polymarket was until recently the largest prediction market by trading volume, but has recently been surpassed by Kalshi. Kalshi and Polymarket are by far the largest real-money prediction markets. Polymarket’s offerings are largely similar to Kalshi’s, but include more markets that I would describe as “ridiculous” if I’m being charitable, or “irresponsible” if I’m not.

Metaculus - the largest prediction *polling* website, which hosts many tournaments, often with prize pools. Metaculus uses a scoring system which rewards accuracy and punishes overconfidence, and maintains rankings for various aspects of forecasting performance. The quality of comments on Metaculus is generally far superior to what you’ll find on other platforms. Metaculus contracts “Pro Forecasters,” selected from the highest performers among its users and offers various forecasting-related services to other organizations. The platform is completely free to

use - my understanding is that it's primarily funded by grant money. For the past few years, Metaculus has also been benchmarking AI forecasting bots against human forecasters to track progress in automated forecasting.

Manifold - a *play money* prediction market, which was very popular for a few years and maintains a devoted following, though it seems to be declining in popularity and relevance these days. Users begin with a small amount of 'Mana' (and can purchase more), which they can use to trade in markets. Questions are user-created and user-resolved, and as a result, the diversity of questions is higher on Manifold than on other platforms. However, this also leads to problems like low-liquidity (many markets only have a few traders), controversial resolutions, and redundant markets. Still, the platform is useful for many niche questions, and is free to use, making it a good starting point to practice prediction market trading.

PredictIt - a prediction market specializing in US politics and elections, which prior to the rise of Kalshi and Polymarket was the place to go for markets on those topics. Those platforms have diminished its relevance, but given its long history and narrow focus, I expect it to stick around.

Good Judgement Project - GJP hosts a prediction polling site (Good Judgment Open) which it uses to identify and certify 'Superforecasters.' It provides forecasting consulting services and offers training to improve forecasting skills. GJP was a pioneer in forecasting *research* and its founder Phil Tetlock is the man behind much of the foundational research on how to make good forecasts.

What is forecasting good for?

Understanding the news

Breaking news events are a popular source of questions for crowd forecasting platforms. The forecasts they produce provide a helpful interpretive layer on top of the traditional news media, taking vague or ambiguous terms like ‘maybe,’ ‘probably,’ ‘unlikely,’ or ‘could’ which are ubiquitous in news reporting and converting them to reasonable probability estimates.

As I write this, the top headline in the New York Times is “*Trump Administration Authorizes Covert C.I.A. Action in Venezuela*,” an article that states “*The authorization is the latest step in the Trump administration’s intensifying pressure campaign against Venezuela ... American officials have been clear, privately, that the end goal is to drive [Venezuelan president] Mr. Maduro from power.*”

After reading the article, I have a few natural follow-up questions: is the US going to get into an armed conflict with Venezuela? How likely is it that Maduro will actually be driven from power? This is precisely the type of question that crowd forecasting platforms help us answer. Polymarket has a “Maduro out in 2025?” market which is currently trading at 18%. On Metaculus, the question “Will the United States attack Venezuela before 2026?” has a community forecast of 34%.

Unlike some of the most ardent proponents, I don’t think it’s right to think of crowd forecasting as a *replacement* for traditional media. Forecasting platforms rely heavily on the news to source questions, and any skilled forecaster will certainly be drawing from news reporting as they formulate predictions. In fact, every major forecasting platform uses mainstream news organizations as the

‘credible sources’ that can be referenced to resolve a question one way or the other.

More fundamentally, I don’t think the primary value of news media comes from reporters and columnists prognosticating about possible futures. News publications still function as a sort of ‘public record’ - if they report something as fact, you can be pretty darn sure that thing actually happened. This is immensely valuable, but it’s also the reason they so frequently use noncommittal qualifiers like ‘might’ and ‘could.’ A news organization won’t report that a thing has definitively happened unless they feel they have credible firsthand proof that the thing has happened. For high profile stories, this amounts to them having at least 99% confidence that the thing has occurred.

This is both a reason that news reporting is still important and a reason why forecasting platforms add value to the news landscape. The news is our best source for “this thing happened” and “this thing didn’t happen,” but it struggles to convey the in-between states: “might happen,” “could happen,” “probably won’t happen,” or the classic, “increasingly likely to happen.” Forecasting platforms do an excellent job adding precision to these in-between states, telling us if that “probably” represents a 55% chance or a 95% chance.

Planning for the future

Beyond interpreting short-term news events, forecasting can be used to anticipate events far into the future. This type of forecasting is useful for organizations that need to plan years in advance: government institutions with multi-year budget cycles, investors trying to anticipate long term market risks and

opportunities, or philanthropists looking to anticipate and mitigate problems that are massive in scale, but slow to develop.

Long-term forecasting helps individuals and institutions *make better decisions*. Many of our decisions are going to be made no matter what - how to allocate the budget, which candidate to vote for, which cause to donate to... we can either guess randomly and hope for the best, or we can *try* to figure out which choices are more likely to get us the outcomes we want.

Say you're an investment firm considering various artificial intelligence-related investments. You'd like to understand how rapidly AI capabilities are likely to progress over the next 1, 5, or 10 years. Around what time do we expect the next few generations of models to be released? How quickly will advances in digital applications transfer to robotics? Are Chinese LLMs likely to surpass American LLMs in the next few years? These are challenging questions for any one person to forecast, and it's easy to succumb to the hype, be needlessly skeptical, or just take a guess that 'feels right' but isn't grounded in anything. Crowd forecasting platforms pose many questions just like these, giving reasonable estimates that can be a useful starting point for investors who lack domain expertise.

Or for a personal example, say that you're a US voter who cares deeply about foreign policy. In particular, you wish to see the Russia-Ukraine war and the Israel-Gaza wars end as soon as possible. In the lead-up to the 2024 presidential election, you likely wondered which candidate was most likely to bring about an end to these conflicts more quickly, Trump or Harris?

Metaculus posed a number of conditional question pairs (e.g. If Trump wins, will the Russia-Ukraine war end by 2026? / If Harris wins, will the Russia-Ukraine war end by 2026?) examining the likelihood of various events given different election results. These

forecasts were a useful tool for gauging which outcomes were most dependent (or not) on who won the election, and thus for understanding which issues should factor most strongly into a voting decision.

Personal benefits

In addition to these *consumer* benefits of crowd forecasts, I also argue that *participating* in crowd forecasting initiatives, that is to say producing and evaluating your own forecasts, can be personally enriching.

Forecasting gives you direct feedback on how clearly you see the world. As the economist Alex Tabarrok put it, staking reputation, credentials, or money on your opinions imposes a “tax on bullshit,” penalizing delusion and rewarding insight.

If you’re catastrophizing about what your least-preferred political party is going to do when they take power, making forecasts about what specific actions you’re sure they’ll take and evaluating those forecasts at a set date in the future can show you if your fears were justified or overblown. If you’re confident that AI is about to rapidly change society, try writing down some concrete impacts you expect to see one or two years from now, and see if you’re correct. If you make this a habit, writing down what you think is going to happen and checking back later to see if you were right, you’ll start to see where your biases and blindspots are, or the topics where you’re especially prescient.

Forecasting also encourages an active, directed style of news consumption that is more conducive to *real learning*, than the passive news reading that most people practice. As you work to produce a forecast for a question, you’ll naturally come up with a bunch of questions that factor into your forecast ... What has

Trump actually said about the Ukraine conflict? How long do modern wars typically last? Have other Republican members of Congress deviated from Trump on the issue? ... and you'll actively seek out the answers to those questions. When you read an article on the subject, you'll read it with a specific frame in mind. How does this new development affect the likely end date for the war? This is very different from scrolling mindlessly (or angrily) through news headlines and reading articles without any idea of what you're trying to get from them.

Personally, I've found that I have a much deeper understanding of topics I've made forecasts on compared to topics where I haven't (even when I've passively read lots of articles about the latter). I largely attribute this to the active style of news consumption that forecasting requires.

Finally, a strong forecasting track record is a way to establish professional credibility on any subject. You succeed as a forecaster by being trustworthy (people can rely on what you say) and insightful (what you say improves upon conventional wisdom). These are prized characteristics across many domains - consulting, investing, management, ... any role that requires planning for the future.

Your first forecast

To this point, Part I has explained what crowd forecasting is and why it's useful. The remainder of this section and all of Part II are about how to make your own forecasts and how to improve the quality of those forecasts.

Here, I'll walk through the basic mechanics of forecasting on each of the two types of crowd forecasting platform. In Part II, I'll cover a variety of concepts, skills, and strategies that will help you make better forecasts.

Interpreting the question

No matter the platform you're using, you'll be presented with a short, 'headline' version of a question, as well as a more detailed explanation (the 'fine print' or 'resolution criteria') that spells out the specific conditions under which the question will resolve Yes or No.³ The detailed explanation *should* cover any foreseeable outcomes, and avoid creating scenarios where it's ambiguous which way the question should resolve. It will also list any sources that will be referenced to resolve the question.

It's very important to read the resolution criteria carefully. Often, a subtle detail that isn't apparent from the question headline can substantially change the odds of each outcome.

³ On prediction markets, all questions are posed as Yes/No. Polling platforms like Metaculus support additional question types such as multi-choice (2+ options where the probability should sum to 100% across all the choices), and continuous probability distributions.

Will there be a bilateral ceasefire in the Russo-Ukraine conflict before 2026?



97 79 comments Closes Dec 31, 2025 1.9k forecasters

Resolution Criteria

This question will resolve as **Yes** if a bilateral ceasefire has gone into effect and stood for 30 days, beginning at any point between January 1, 2025 to December 31, 2025, inclusive. A ceasefire is bilateral if it applies to the majorities of combatants on both the Russian and Ukrainian sides, respectively. A ceasefire is deemed to have stood as long as no reliable sources report that the ceasefire has broken down or is no longer effective. If no ceasefire has stood for 30 days before January 30, 2026, this question will resolve as **No**.

The ceasefire must apply to all military operations in all official and disputed Ukrainian and Russian territory. In other words, a limited ceasefire (such as granting safety to humanitarian corridors or specific regions) is insufficient to resolve the question as **Yes**.

Examples of ‘headline’ and ‘fine print’ versions of a question from Metaculus. Always read the resolution criteria carefully. It’s not uncommon for a small detail to significantly impact the likely resolution for a question

Question writers will occasionally miss an edge case, or accidentally phrase part of the description so that some possible outcomes will have an ambiguous resolution. If you notice this, you have a few options:

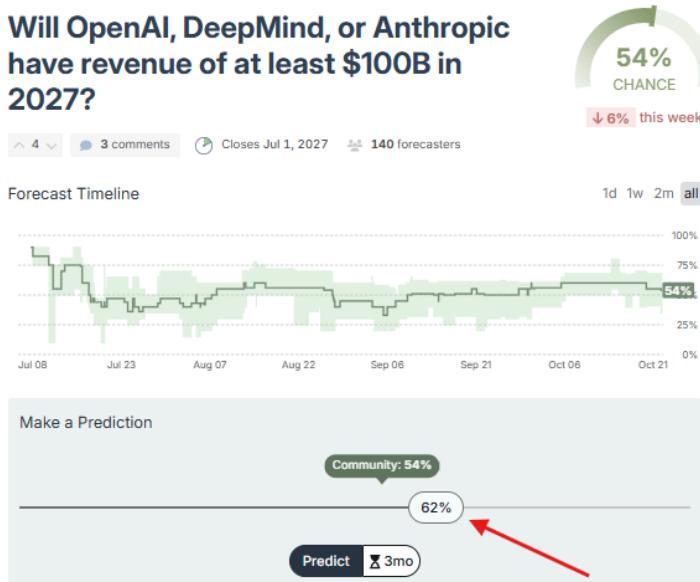
- Leave a comment asking for clarification and hold off on participating in that question until it’s given
- Participate anyway, but factor the possible ambiguous outcomes into your forecast. Determine which resolution is more likely under the ambiguous outcome, but do so knowing that you might be incorrect
- Avoid the question entirely

What you should *not* do is ignore the ambiguity and then complain in the comments after the fact when the edge case occurs and the question resolves in a way you didn’t expect.

Forecasting on a prediction polling platform

Metaculus is a good choice for anyone dipping their toes in the waters of forecasting. It's the largest prediction polling platform, beginner friendly, and is simple to get started on. To make a prediction:

1. Pick a question that looks interesting
2. Do some background research so you know what it's asking
3. Determine the probability you think is correct
4. Input that probability directly and submit
5. As you learn more, as time passes, or if you change your mind, update or withdraw your prediction (optional but recommended)



To forecast in a prediction poll, just directly input the probability you think is correct, and update whenever your estimate changes

Once the outcome is known, moderators will *resolve* the question to Yes or No (in rare cases they may ‘annul’ the question early or resolve it as ‘ambiguous’). You’ll achieve a good score if your prediction was more confident in the eventual outcome (or less confident in the wrong outcome) than the community average, for an extended period of time. You’ll improve your ranking on the site by consistently achieving good scores across a large number of questions.

Forecasting on a prediction market

Prediction markets are slightly more complicated to use, but the ‘scoring’ is very simple. Prediction markets use an *order book* to coordinate the purchase and sale of ‘shares’ of Yes or No outcomes. To participate in the market, you can

- a. buy or sell shares to fulfill orders already listed in the order book
- b. place a *limit order* that will appear in the order book, where you agree to buy or sell a set number of shares at a set price, should another trader take you up on the offer. This order will stay in the order book until somebody fills it or you decide to cancel it.

If you feel that the market price is too low for a particular outcome on a question, you buy shares of that outcome. Exactly how many shares you *should* buy is a complicated question, and one I’ll cover in more detail later. As a basic rule, you should invest more heavily the more mispriced you think the market is, and should avoid investing too large a fraction of your bankroll in any one market.



Will Trump eliminate the Department of Education this year?

2 413 ↑ +

4.2% chance ▼ 10.8 ◎

Kalshi

Order book ◎ 💚 Rewards

Trade Yes Trade No Graph

	Price	Contracts	Total
	5.4¢	600	\$2,378
	5¢	2,500	\$2,378
	4.8¢	15	\$2,253
Asks	4.2¢	53,622	
	Trade Yes Last 4.2¢	Trade No	Rewards
Bids	4.1¢	53,864	\$2,208
	4¢	28,722	\$3,357
	3.9¢	2,015	\$3,436
① ⚙ + 0.1¢	3.8¢	1,000	\$3,474

Example of a prediction market currently listed on Kalshi. ‘Yes’ shares can be purchased for \$0.042, or sold for \$0.041 (likewise, ‘No’ shares are trading at \$0.958).

At the end of the year, each share will be redeemable for \$0 or \$1, depending on the outcome.

Once you own shares of an outcome, you can hold them until the question resolves or sell them if you feel the current market price has corrected to a reasonable (or *overpriced*) number. When a question resolves to Yes or No, shares of the corresponding outcome cash out for \$1, while shares of the wrong outcome become worthless.

You succeed in a prediction market by growing your bankroll. If you are consistently able to judge when markets are improperly priced and buy or sell shares at those prices, you’ll gain money over time. Otherwise, you’ll gradually (or quickly) lose your money.

II. Strategies and Skills

How to make better forecasts

Practice

The only way to get better at forecasting is to practice forecasting. Forecasting is a skill like any other - if you devote consistent effort to it over a long period of time, you will get better at it. If you take one thing away from this guide, let it be this: *forecasting is a skill you can improve at, but there is no way to improve at it that doesn't require doing it a bunch of times and learning from your mistakes.* To practice forecasting:

1. Find or create a list of questions about the future. Try to keep most of them short-term (a few weeks to a few months) so that you'll get quicker feedback. Each of the forecasting platforms I've mentioned has a huge library of questions you can practice with. Metaculus runs free tournaments throughout the year with short timeline questions that are an excellent way to practice. You can also come up with your own questions, about any topics you find interesting. Make sure there's a clear Yes/No outcome and a defined end date.
2. Produce a forecast for each question. Always make a probabilistic forecast, e.g. "there's a 75% chance this will happen." Use whatever method you'd like to come up with your predictions. I suggest starting quick and simple and then gradually incorporating some of the other strategies discussed here as you go. Record each prediction somewhere along with the date you made it.
3. Once the outcome of a question is determined, review your prediction and grade it based on the resolution. I'll discuss scoring methods in a moment, but as a starting point, did your most confident predictions resolve the way you expected? Were you overconfident or underconfident about anything? Are there certain topics you seem to do better at?

That's all it takes - find a question, make a forecast, track the result.

Try to make this a habit you incorporate into some aspect of your daily routine. If you read the news each morning, make one or two forecasts about an event that's in the news. Join a free forecasting tournament and try to make at least one prediction on each of the questions. When someone on a podcast or YouTube video makes a strong claim about something they think will happen, take a minute to consider what *you* think, and try making a probabilistic forecast about it.

Developing forecasting skill takes time. The events we care about often take weeks or months to play out, so there's usually a delay between when we make our predictions and when we get feedback on their accuracy. But with consistent practice, by making many forecasts and reviewing our results, we can gradually hone this skill.

Measuring success

Once we've established a habit of writing down our predictions, how do we figure out if we're improving over time? Earlier I discussed trustworthiness and insightfulness as two properties of good forecasts. In this section, I'll cover methods to quantify these attributes and discuss how you can use them to track your progress.

Calibration plots

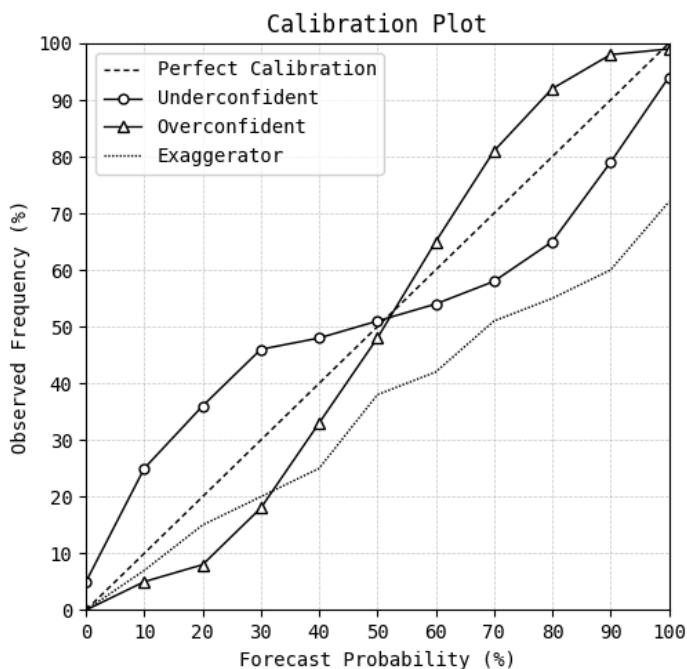
A calibration plot (sometimes called a reliability diagram) shows how trustworthy or ‘well-calibrated’ a forecaster is. To determine a forecaster’s calibration, we need to look at how often things *actually occur* when the forecaster assigns a given probability to that thing. A calibration plot shows this visually - the x-axis represents probabilities given by the forecaster, while the y-axis shows the *observed rate* of occurrence.

If a forecaster is perfectly calibrated, the plot will show a straight diagonal line from the bottom left to the top right. Points below this diagonal indicate that things happen less frequently than the forecaster predicts they should, while points above the line indicate that the forecaster understates their chances. A forecaster might exhibit *overconfidence* where they give too many forecasts near either extreme (0% or 100%), or *underconfidence*, where their forecasts stick too close to the middle (50%).

Calibration plots are easy to create and an easy way to track your own calibration. To make one:

1. Maintain a spreadsheet with every prediction you make, including your forecasted probability and what the eventual outcome was

2. Group your forecasts into buckets, for example 0-10%, 10-20%, ... 90-100%. Start with fewer, wider buckets and increase the number of buckets over time as you accumulate more forecasts
3. For each bucket, collect every forecast you've made that falls within that bucket. Compute the % of the time questions in that bucket resolved to Yes for the forecasts in that range.
4. Plot the observed probability for forecasts in each bucket, along with the ‘perfect calibration’ diagonal as a reference.



Example of a calibration plot, showing how often things actually occur when a forecaster assigns them some chance of happening. An *overconfident* forecaster makes too many forecasts close to 0% and 100%. An *underconfident* forecaster makes too many forecasts near 50%, and would do well to ‘extremize’ their predictions. The ‘exaggerator’ systematically overestimates the chances that things will happen.

As you build up a history of forecasts, you should start to get an idea of how well your predictions reflect the true probability of an outcome. If you find you're consistently underconfident or overconfident, you can take this into consideration as you make subsequent forecasts.

Note that you'll need many forecasts to get an accurate picture of your calibration. For 5% confidence predictions, you would only expect to see one Yes resolution out of every 20 forecasts. The more extreme the forecasts are, the more samples you need in order to determine your calibration. Starting with large buckets and decreasing the bucket size over time is a good way to get use out of a calibration plot early on.

Trade outcome plots

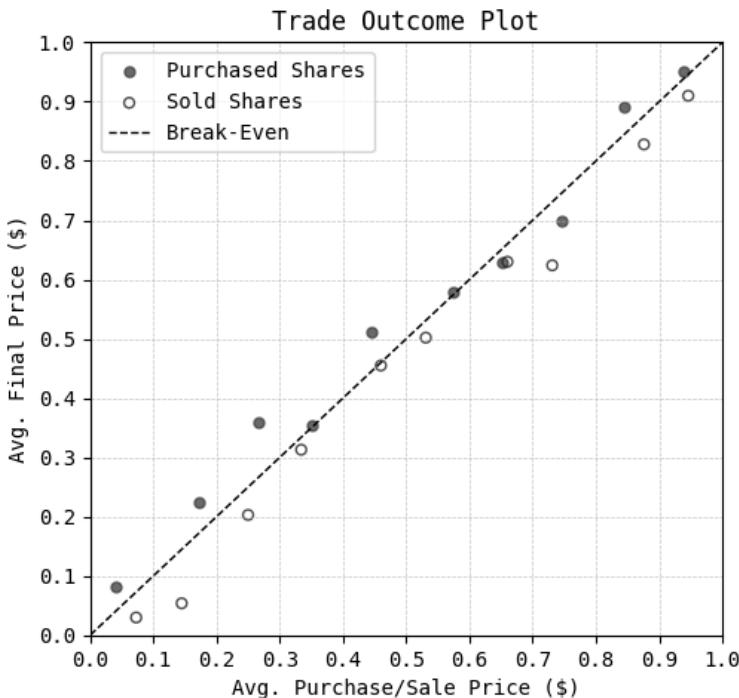
Calibration plots are easiest to use with prediction polling forecasts, where you always provide a specific probability for a given question. You certainly *can* do this when trading in prediction markets. In fact, determining what you believe is the ‘fair price’ for a market is recommended and helpful for determining how much of your bankroll to spend on a market. However, a variant of the standard calibration plot which I call a ‘trade outcome plot’ is useful for tracking your prediction market trading results.

You make money in a prediction market when the shares you purchase at a given price redeem for \$1 *more often than the probability implied by that price*. For example, if you look at all the shares you've ever bought for \$0.25 and find that more than 25% of those shares were eventually redeemed for \$1, you've made money on those purchases.

We can construct a plot that shows exactly this - for different price ranges, what is the average ‘final price’ (\$0 or \$1) that shares we purchased in that price range were sold for. For each price range, we should look for two numbers:

- The average final price for shares we *bought* in that range
- The average final price for shares we *sold* in that range

If we’re making good trades, the average final price for shares we buy should be *above* the purchase price, and the average final price for shares we sell should be *below* the sale price. By comput-



Example of a trade outcome plot, showing the average final price of shares purchased and sold across different price ranges. Transactions are grouped in \$0.10 buckets. For purchased shares, we want points above the diagonal, which indicate that shares

were eventually worth more than the purchase price, on average. For sold shares, we want points under the diagonal, which indicate that shares were eventually worth less than the price they were sold for.

-ing and plotting these for each price range, we can see how skillful our trading is across each range.

Though it looks similar, we interpret this plot differently from a calibration plot. With a calibration plot, we strive for values *along the diagonal*, which indicate perfect calibration. With a trade outcome plot, points along the diagonal indicate ‘neutral’ trades. We neither gained nor lost money from them, on average. The farther *above/below the diagonal* (for buys/sells) our datapoint is, the more profit we’ve made on trades in that range. While trade outcome plots do hint at calibration, they mainly convey *insightfulness*, by showing how our trades outperform or underperform the market.

Brier scores

A *Brier score* gives us an absolute measure of how ‘correct’ a forecast was, given the outcome. The Brier score formula is:

$$\text{Brier score} = \frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2$$

f_t is the forecast probability (0 to 1)

o_t is the outcome (0 or 1)

N is the number of forecasts being evaluated

For a single prediction, the Brier score is just the difference between our forecast and the result, squared. For a group of questions, we take the average of the individual Brier scores.

For any pair of forecasts, the one that predicted a higher probability of the true outcome will have a lower Brier score (Brier scores measure the prediction *error*, so lower is better). A perfect forecast has a Brier score of zero, whereas a perfectly *bad* forecast has a Brier score of 1.

Brier scores have limited utility for evaluating forecasts on their own, because the types of questions we forecast can have widely varying *difficulty*. For example, it's much more difficult to predict which baseball team will win a game than it is to predict if the sun will rise tomorrow. I can achieve a perfect Brier score by always predicting a 100% chance that the sun will rise. Achieving such a low Brier score for baseball game prediction is impossible.

Brier scores *can* be used as a measure of forecast quality, but only when comparing predictions on questions of equal difficulty. You could use Brier scores to track improvements to, say, a weather forecasting model, because the underlying difficulty of weather prediction doesn't vary year-to-year.

Skill scores

To judge forecasts, it's often more useful to look at *skill* scores. Skill scores compare an absolute measure, such as the Brier score, to a *baseline*. They tell us how good a forecast is, relative to that baseline. The baseline could be random guessing, a simple rule of thumb (for example our climatology prediction in the case of weather forecasting), or the median of all the individual forecasts in a prediction poll. One such score is the *Brier skill score*:

$$\text{Brier skill score} = 1 - \frac{\text{Brier}_{\text{you}}}{\text{Brier}_{\text{ref}}}$$

where $\text{Brier}_{\text{you}}$ is your personal Brier score

and $Brier_{ref}$ is the Brier score of the baseline forecast

Here, higher numbers are better. Positive numbers indicate that a forecast outperformed the baseline. Negative numbers indicate that a forecast was worse than the baseline.

Prediction polling sites all use some form of skill score as the basis for their scoring system, typically comparing each individual forecast to the median of all forecasts on a question, or using a comparable method. This makes it possible to compare forecasters who answer different questions, with varying difficulty. Your score for each question is based on how good your forecast was, *relative to the other forecasts on that question*.

Log scores

Another scoring rule is the *log score*. Metaculus scoring is based on the log score.

$$\text{Log score} = \ln(p)$$

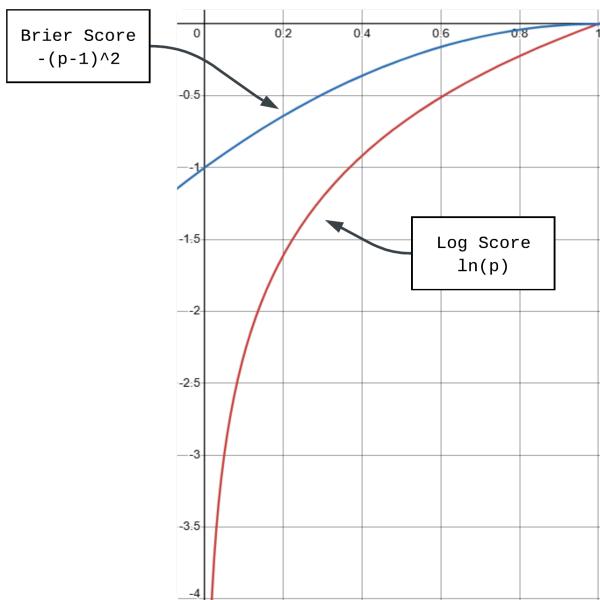
Where p is your forecast probability for the outcome that occurs (0 to 1)

With this scoring rule, higher is better, and the highest score you can achieve is 0. The log score can also be turned into a skill score by comparing it to a reference.

$$\text{Relative log score} = \ln(p) - \ln(\text{ref})$$

The major difference under log scoring is that your score can be *infinitely bad*. If you predict a 0% chance of an outcome and it happens, your log score is $-\infty$ (Metaculus limits your prediction to 0.1 and 99.9 to prevent this). Overconfidence is penalized more

harshly with log scores than with Brier scores. If you try tournament forecasting, make sure you understand what scoring method the tournament host is using. If log scores are used, it's worth being extra cautious when making extreme predictions. Err on the side of lower confidence, or devote extra research time before making such predictions. A single very bad prediction can ruin your score for an entire tournament if you aren't careful.



Graphical comparison of log scores vs. Brier scores. The Brier scores are inverted so that higher is better, for simpler comparison to the log scores. Brier scores are more forgiving of overconfidence. If your forecast probability (p) is zero for an outcome that occurs, the worst your Brier score can be is 1, whereas your log score drops to $-\infty$.

Writing

Perhaps the single most effective habit you can develop to improve at forecasting is *writing*. Write about what the question is asking. Write about what you found in your research and how you arrived at the probability you arrived at. Write about why you were right or wrong once the outcome is known. Write about what you learned that you'll apply to your next prediction.

Writing forces us to organize our thoughts and make them legible. The moment we try to write about something, we realize that our grasp of the subject isn't quite as firm as we believed it was. We notice gaps in our knowledge, flaws in our reasoning that weren't apparent from merely *thinking* about the topic but which become obvious when we try to put those thoughts into words.

This makes writing difficult, slow, and taxing, but it's also what makes it so useful in forecasting. Those flaws in our reasoning are there whether we're aware of them or not. Writing increases our chances of catching and correcting them before they wreck our predictions.

Writing to understand the question

Before I make a prediction, I try to re-state the question in my own words, as though I'm explaining it to a smart friend who has little or no background knowledge about the topic. This forces me to look closely at the description and resolution criteria, and typically requires doing basic background research on the question. Explaining it in simple terms for a general audience takes a surprisingly in-depth understanding of the topic, so if I'm able to do that, I'm probably in good shape to make a prediction.

The act of writing the summary (as opposed to just doing the reading) is critical. I'm sure you're familiar with the experience of reading an article or a book, only to discover that you've retained almost nothing when you try to discuss it a day later. Writing is a solution to that problem. As you write, you're forced to refer back to your sources, review the things you've forgotten, and seek out bits of information you realize are missing from your explanation.

Here's an example of me doing this for a question posed in a 2023 Metaculus tournament:

Will the oil transfer from the FSO Safer finish in 19 days?

My write-up: *"The FSO Safer is an oil tanker that was abandoned in the Red Sea off the coast of Yemen back in 2015, containing over 1 million barrels of oil. It has remained there ever since amid regional conflict and disputes over who owns the oil and ship."*

In the years since the ship was abandoned, its hull started corroding, and the inert gas that typically fills the empty space in the oil tanks to prevent the oil from exploding has gradually leaked out. This has led to increasing concern about a massive oil spill and/or explosion, which would shut down nearby ports and ravage the local fishing industry.

In 2022, the UN launched a crowdfunding campaign to raise funds for a salvage operation to offload the oil from the vessel. The UN purchased the vessel earlier this year, and the salvage operation began in July. The salvage company, Boskalis, stated that the offloading of the oil was expected to take 2-3 weeks. This question opened around July 25th, the date that the oil transfer from the FSO Safer to the salvage vessel began. The question is asking whether the transfer will finish within the anticipated timeline."

Writing to make a coherent forecast

Once you've re-stated the question and context, you can apply the same process to your prediction. Explain what factors you're considering, assumptions you've made, key findings from your research, any quantitative modeling you've done... forecasting approaches will vary widely for different types of questions, so the exact details will vary, but in all cases you should be able to describe the basic process you used to arrive at your number.

Again, pretend you're talking to a smart friend who is unfamiliar with the subject. Where will they be confused and ask follow up questions? What parts of your argument will they find unconvincing? If you can write a few paragraphs coherently laying out your reasoning, you've probably done a decent job researching the topic and considering possible outcomes.

Here's what this looked like for the *FSO Safer* question from the previous section:

"I forecasted a high chance of this happening, opening at 90%:

- *I expect the salvage company would err on the side of being too conservative with their timeline, rather than too aggressive*
- *The timeline for transferring the oil seems easy to predict, relative to other parts of the operation. You know how much oil there is to pump and how fast your pumps go.*
- *When the question opened, pumping had already started. So by this point, many of the more uncertain parts of the process were already finished*
- *The salvage company has been very transparent about the whole operation and things seem to be going very smoothly.*
- *I've raised my estimate a bit after we've received a few early updates about the progress of the transfer. It took less than a week for the first*

half of the oil to be transferred, so things appear to be very much on track.



Chart showing oil transfer progress for the FSO Safer salvage operation, vs. pace required to finish by question deadline

The last few updates have made me slightly concerned as the pace appears to be slowing down, but it's hard to tell if this is real or if the timing of each progress report is just off a bit. I'm just going off of periodic Tweets from a few accounts tracking the operation, and the exact timing of each update could be off by up to about a day in either direction.

For now I've lowered my estimate slightly, but the next few days should help clarify if there's actually been a slowdown or not."

Writing to remember what you were thinking

The previous two sections described how writing can be helpful for improving predictions, as you're making them. But writing is also a tool for learning from successes and mistakes, *after a question has resolved*.

A major challenge with forecasting is that there can be a long delay between when we make a prediction and when we get

feedback on whether that prediction was good or bad. Describing in writing what we’re thinking about at the moment we make a prediction is a way to overcome this challenge. Once a question resolves (or really at any time after we’ve made a prediction), we can revisit the research, assumptions, and logic that went into our forecast, and see how well they held up against reality.

If we don’t have a written record to check, this is impossible to do reliably. Not only will we simply forget much of what we were thinking, we’ll also inevitably misremember the reasons we made the prediction, often in a way that makes our former self look wiser than they really were. This comforts us in the moment, but diminishes our ability to learn from our mistakes and improve over time.

To learn from our past predictions, it’s critical that whenever we write about a prediction, we do so *as honestly as possible*. If you didn’t put much effort into a question, write that down! If you tried to make an objective forecast about a question where you strongly preferred one outcome, and worried that your preference was clouding your judgement, write that down! If you were testing out a new statistical modeling technique and weren’t sure if you applied it correctly, write that down! If you did extremely thorough research and feel supremely confident in your forecast, write that down!

Ultimately, our goal shouldn’t be to make our logic sound impressive or to feel smart, our goal should be to *get better at forecasting over time*. The best way to do that is to honestly describe what we’re thinking and feeling when we make a forecast. If we do this, we can look back on our earlier forecasts and get a clear picture of where our logic was flawed (or flawless), and apply what we learn to our next round of predictions.

Here's an example of a write-up I did for a question about Javier Milei's vote share in a 2023 Argentinian primary election:

"To get an estimate here, I'm just doing a very simple polling average, averaging together all the recent polls, and distributing all the undecideds proportionally to each party. To get 25th and 75th percentiles, I'm honestly just picking numbers that look about right. I'd like to read more about good strategies for averaging polls and getting reasonable margins of error, but for now I'm just going for something in the right ballpark."

The distribution I submitted for the question randomly happened to be slightly wider than the community average, and thus gave higher probabilities for 'extreme' outcomes. Javier Milei ended up performing much better than expectations with his vote share falling into one of those extremes, so I got a great score on the question. Without my write-up, it would have been easy to think, "wow, I crushed that question, I must be great at forecasting elections!" But if I look back at my 'reasoning' I have to acknowledge that this result was just luck.

Base rates

When we first start researching a question, it can be difficult to get even a rough sense of how likely each outcome is. Is it closer to 20%, 50%, or 80%? Or more like 1% or 99%? An effective way to figure out a reasonable ballpark estimate for our forecast is by finding a relevant *base rate*.

A base rate for an event is *the rate at which similar events have occurred in the past*. I'll illustrate this with several examples.

Will it rain in Boise on November 1st?

Base rate: Percentage of years over the last 100 years in which it has rained on November 1st in Boise

Will a Scottish independence referendum be called before 2027?

Base rate: Percentage of years over the last 50 years in which a Scottish referendum has been called.

Will the government shutdown (beginning October 1st) end before October 15th?

Base rate: Percentage of past government shutdowns that have lasted at least 15 days

Will India's Chandrayaan-3 mission successfully land a rover on the moon?

Base Rate: Percentage of past attempts to land rovers on the moon that were successful

Base rates give us a reasonable starting point to build our forecast on. Particularly for an event where there are many historical analogues, base rates give us an estimate of how likely that event is *in general* which is usually a decent guess for how likely that event is *this time*.

Current circumstances will (and should) prompt us to deviate somewhat from the base rate. For many forecasters however, *overreacting* to current events tends to be a more common issue than *underreacting* to them. In my experience, directly using a base rate as my forecast has often been enough to outperform the community average on crowd forecasting platforms like Metaculus.

There are usually several possible base rates we could use for a question. Consider this question: *Will Ohio Issue 1, a ballot initiative which would increase the vote threshold required for future amendments to pass, be approved in the August 8, 2023 election?* Some possible base rates for this question are:

- The percentage of ballot initiatives across Ohio's entire history that passed
- The percentage of ballot initiatives *in the last 25 years* that passed
- The percentage of ballot initiatives *asking similar questions to this one* that passed in *any state* over the last 25 years
- The percentage of ballot initiatives *asking similar questions to this one* that passed *in Ohio* over the last 25 years

And surely many other variants of these as well. So which do we choose? There's no universal rule I can give, and ultimately it comes down to using your best judgement about which rate (or which set of rates - you can pick a few and average them) is most relevant to the specific question you're researching.

In cases like this one, there's a tradeoff between *sample size* and *relevance*. Looking at every ballot initiative across Ohio's entire history gives us a large sample, but ignores the contents of those initiatives, and also includes votes from very different political eras. Looking at ballot initiatives that asked similar questions to

the one in question, across *all states* gives us a sample more relevant to the current vote, but includes examples from non-Ohio electorates. Narrowing further to only include similar ballot initiatives *in Ohio* would give us the most relevant sample of all, but unfortunately this reduces our sample size to zero.

In this case, I decided to consider ballot initiatives from any state, that asked a similar question to the one asked by Ohio Issue 1, from the last 10 years. This resulted in a sample of 12 initiatives , roughly 20% of which passed.

The goal of finding base rates isn't to give you a perfect answer (though it may well give you a good one), it's to give you a reasonable starting point to build from. When I first began researching the Ohio Issue 1 question, I had no sense at all for how likely its passage was. The base rate told me I should be starting from a probability around 20%, only deviating from that if there was strong evidence to suggest otherwise.

Minimum viable models

Different types of questions require different ‘styles’ of forecasting. On one end of the spectrum we have *model-based* or statistical forecasting, where our subjective view doesn’t factor into our prediction, beyond the choice of what model and what data to use. On the other end, we have *intuition-based* forecasting, where our forecast is almost entirely based on our subjective ‘feel’ for the question. I’ll discuss these two approaches in this section and the next.

Statistical modeling is especially useful for questions that are rich in data or which involve repeated trials. Some examples of this are elections, disease spread, financial markets, sports, and physical phenomena like weather. Contrast this to a question like “Will there be a cease-fire in the Russia-Ukraine war before 2026?” which lacks obvious data to base a model on, and which is highly dependent on decisions made by specific individuals.

Specific statistical modeling methods are beyond the scope of this guide, but I do want to offer a general guideline for questions that require it: always start with the *simplest possible model that’s at least a little bit useful*. I call this the ‘minimum viable model’ approach, and it helps us deal with some of the challenges of forecasting in competitive settings like tournaments or prediction markets.

First, unless you’re specializing in one narrow domain, you’ll likely be making forecasts on questions across a wide range of topics. Simple models are important because *you don’t have time to build a complex model for every question*. You may have heard of the ‘80/20 principle,’ the observation that for many tasks you can achieve 80% of the results with 20% of the effort. Starting with the minimum viable model is how we apply this principle to

forecasting. In tournaments or prediction markets, many participants won't build *any model at all*. Building a simple model will get you ahead of much of the field, while helping you avoid spending too much time on any one question, neglecting the rest.

Second, simple models help mitigate the problem of *overfitting*. Overfitting is a problem in statistical modeling where your model is too complex for the available data. An overfit model will often perfectly explain every previous data point, but fail to extrapolate to new data.

Say you're trying to predict the outcome of a coin flip. You've observed 10 flips of the coin: H-T-H-H-T-T-T-T-H-T. Consider the following two models:

Model 1: Take the average rate of heads across all previous trials. Use that probability for the next flip (in this case 40% heads, 60% tails)

Model 2: Always predict 100% chance of heads for the first flip. Following a single heads, give 67% probability to tails and 33% probability to heads. Following a single tails, give a 50% chance for each outcome. Two consecutive heads is followed by tails 100% of the time. If we see two or three tails in a row, the next flip is tails 100% of the time, but if we see *four* tails in a row, the next flip is *heads* 100% of the time.

Applied to our sample data, Model 2 correctly predicts the observed results far more confidently than Model 1, but Model 1 will clearly do a better job when applied to any large sample of new data. Model 2 is *overfit* to our dataset. It tries to explain every single data point perfectly, but as a result sees patterns where there is only noise.

	Complex Model - Historical Data											
% Chance Heads	100%	33%	50%	33%	0%	50%	0%	0%	100%	33%		
% Chance Tails	0%	67%	50%	67%	100%	50%	100%	100%	0%	67%		
Flip Result	H	T	H	H	T	T	T	T	H	T		
Brier Score	0.00	0.11	0.25	0.45	0.00	0.25	0.00	0.00	0.00	0.11		
Avg. Brier Score												0.12

	Simple Model - Historical Data											
% Chance Heads	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%
% Chance Tails	60%	60%	60%	60%	60%	60%	60%	60%	60%	60%	60%	60%
Flip Result	H	T	H	H	T	T	T	T	H	T		
Brier Score	0.36	0.16	0.36	0.36	0.16	0.16	0.16	0.16	0.36	0.16		
Avg. Brier Score												0.24

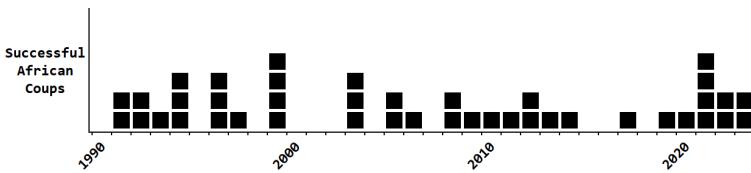
First, we compare how well our two models (described in the text) explain the data we've observed *previously*. The complex model perfectly anticipates half of the flips, and achieves a lower average Brier score than the simple model...

	Complex Model - New Data											
% Chance Heads	100%	33%	0%	0%	0%	0%	50%	0%	33%	50%		
% Chance Tails	0%	67%	100%	100%	100%	100%	50%	100%	67%	50%		
Flip Result	H	H	H	H	H	T	T	H	T	H		
Brier Score	0.00	0.45	1.00	1.00	1.00	0.00	0.25	1.00	0.11	0.25		
Avg. Brier Score												0.51

	Simple Model - New Data											
% Chance Heads	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%	40%
% Chance Tails	60%	60%	60%	60%	60%	60%	60%	60%	60%	60%	60%	60%
Flip Result	H	H	H	H	H	T	T	H	T	H		
Brier Score	0.36	0.36	0.36	0.36	0.36	0.16	0.16	0.36	0.16	0.36		
Avg. Brier Score												0.30

... However, when we use each model to predict *new* data, the complex model performs much worse. Even though the simple model isn't perfect (we know the 'ground truth' probability is 50% for both heads and tails), it avoids the severe overconfidence that the first model suffers from.

This same issue can arise when we attempt to model other real world phenomena and build forecasts. Consider the question “will there be a successful coup in any African country in 2026?” With a bit of research, we find the following data on the incidence of coups in Africa over the last 35 years:



Number of successful coups in African countries each year since 1990. Data from projects.voanews.com/african-coups

We *could* construct a model that perfectly explains every single data point on this plot, just like our model that predicted all the previous coin flips (“following a year with exactly 2 coups, there’s a 100% chance of at least 1 coup the following year”, etc.). Or, we could observe that the pattern looks quite random, and note that there has been a successful coup in ~70% of years over the past few decades, and start our forecast there. The 70% model is likely to do just as well or better than the complex model when applied to new data.

I don’t mean to suggest that you should never add any complexity to your models. Rather, my advice is to always *start* with a simple model and only add complexity if you can be confident you aren’t overfitting to random noise. Our time is often better spent seeking out *more or better data* than on increasing the complexity of our model. A simple model built on excellent, representative data will outperform a fancy model built on poor data.

Intuitive forecasting

For some questions, there's no clear way to apply a statistical model, and even a rudimentary base rate is hard to come by. And even when we do have a basic model, there's still the question of how much we should deviate from the model based on the specific current circumstances. In these cases, we may have no choice but to rely on our intuition. Questions where the outcome relies on individual decisions made by a small number of people generally fall into this category.

This, in my view, is the most challenging aspect of forecasting and the hardest to explain, but also where the greatest opportunity is to get an edge over the crowd. Questions that require intuition are the ones where I most often observe large differences between my personal forecast and the crowd forecast. Intuition-based forecasting is a difficult process to articulate, but I'll do my best to describe how I think about it.

There are basically two parts: first, there's the part where we research a question, build up context, consider different perspectives, and simulate possible futures in our minds. As we do this, we start to develop a vague sense for *which outcomes feel most likely* and also *how confident we feel about our assessment*. Then, there's the part where we try to turn this vague feeling into a numerical probability.

We need to be good at both parts if we want to produce a good forecast. We might feel confident that our team is likely to win its next match. We can correctly interpret that feeling as '90% confidence,' but if we're just delusional about our team's talent, our forecast is still terrible. Alternatively, we might correctly assess that our team is slightly more skilled than the opponent and has an advantage in the match, but if we don't know whether to

interpret that assessment as a 55% chance or an 80% chance, we likewise won't make a strong prediction.

Here's a recent example where I applied purely intuitive forecasting: following the 'indefinite' suspension of Jimmy Kimmel's late night show from ABC, there were prediction markets created on when (and if) his show would return to the air. This is a question where there was not much data to go off of, and which was dependent on decisions made by just a few people.

I started by reading news articles until I felt I had the full context and wasn't missing any key pieces of information about the story. I listened to a few podcasts by media analysts who are subject matter experts on the companies and people involved. At this point, I had a vague sense of what I thought was going to happen. It *felt* to me like:

- The incident in question was minor
- The public backlash against Disney for taking Kimmel off the air was significant and growing
- Both Jimmy Kimmel and Disney wanted to resolve the situation and move on
- The longer the status quo continued, the more severe the consequences for Disney
- The Trump administration had overplayed its hand and had even lost support from many Republicans on the issue

Note that all of these bullet points are partially or completely subjective, derived entirely from the 'sense' I got from reading the news and listening to discussion about the topic. Collectively, these impressions of what was going on led me to think there was around a 75% or 80% chance Kimmel would return to the air within a few weeks. At the time, prediction markets gave this closer

to a 35 or 40% chance. When I saw those odds, they just *felt wrong*. In this instance, my intuition was correct (Kimmel was back on the air within a week) and this ended up as a strong prediction for me.

There are many ways that this process can go wrong, and it has for me on many occasions. This was a particularly positive example. This style of forecasting presents all sorts of opportunities for personal biases to influence and distort our reasoning. It's not uncommon to have missed a critical detail that renders all our other observations irrelevant, or to fixate on something that we later realize was insignificant.

I wish that I could give a more structured approach for dealing with questions like these, but this really is the way I handle them, and it's given me good results. I can offer these suggestions:

- Make sure you do your research before resorting to an intuitive forecast. Sometimes there's a good base rate or data source, but it just requires a bit of extra effort to track down. Only use intuition if there's no better way, or as a final adjustment to a more rigorous forecast
- You'll only get good at this by practicing a lot, with feedback, and in particular by noticing where you tend to go wrong. Whenever you make a forecast, observe how you *feel* making the forecast - how confident/nervous/emotional are you? After the outcome is known, think back to how you felt and try and use the example to calibrate yourself for future predictions
- Read about the many cognitive biases we're susceptible to, try to spot when they're skewing your reasoning, and develop strategies to mitigate them
- Do 'calibration training' such as the one created by Clearer Thinking⁴. This helps with the second part I mentioned earlier, converting your vague sense of likelihood into an actual number

⁴ www.clearerthinking.org/tools/calibrate-your-judgment

- Learn which domains you have good intuition for and which ones you have poor intuition for, and update the confidence of your predictions accordingly

Doing your own research

This might seem obvious, but it needs to be said - you should actually research every question before you make a prediction! Some questions warrant deeper research than others (more on that in a moment), but *all* questions warrant at least a bare minimum.

My sense is that for any given question on a forecasting platform, most participants are half-assing their research (if they're doing any at all) and anchoring heavily to the public community forecast. In a way, this is what you'd expect. Good research can be tedious, and the community prediction is typically quite reasonable, so it makes sense that many people just won't expend the effort required to do independent research. This means that with surprisingly little effort, you can get a leg up on the average forecaster.

It's difficult to give a universally-applicable research strategy since questions vary so much, but I'll offer a few general insights that I've found helpful.

A few sources provide most of the useful information

Something you notice when you read lots of articles about a particular story is that a small number of quotes keep reappearing, and every article is saying almost exactly the same thing, just with slightly different phrasing. This is because there are only a few people with *real information*, information from sources with firsthand knowledge of the situation, and everyone else is just referencing those same nuggets of information. A key to efficient forecasting research is learning to track down this small set of people that supply most of the good information for a given topic, and figuring out how to appropriately parse the things they say.

Earlier I mentioned the Jimmy Kimmel suspension saga. Matthew Belloni is one of the best-connected reporters in Hollywood, and for stories like that one, much of the information that makes its way into the broader media sphere is coming from him (and he is getting it directly from the people involved in the story). Rather than read 50 articles about the story, I found it just as effective to monitor his X account and listen to his podcast. Having looked to him for several different Hollywood-related questions in the past, I now have a decent idea of how to interpret the things he says and apply them to my forecasts. There are people like this for virtually every niche and if you can determine who they are, you can save yourself the effort of combing through dozens of secondhand reports, while still getting the best available information as soon as it becomes public.

Research triage

Learn to quickly assess if a question actually needs deep research or if a quick scan is sufficient. For many questions, even a quick scan is enough to classify them as “less than 1% chance of happening” or “greater than 99% chance of happening.” When you’re new to forecasting, it may not be obvious when this is the case, so err on the side of doing more thorough research and making less confident predictions. You’ll soon start to recognize the questions that don’t warrant much effort, and those that do. This is an important skill because you have limited time and limited energy, and you want to allocate as much of them as possible to questions where there’s a chance to gain a large edge over the field by doing in-depth research.

LLMs

LLMs are a great research tool. Specifically, LLMs trained to utilize web search (ChatGPT 03 or better, or its non-OpenAI equivalents) which can quickly conduct a large number of searches, sift through dozens of sources, and then provide summaries and links to the source material. It would be silly *not* to take advantage of that. However, realize that this is the bare minimum. Assume that all of your competition has also taken the requisite 10 seconds to ask the LLM about the question. (And of course, all the usual LLM caution applies, check the source material, hallucinations happen, etc.).

Beware ceding all your thoughts to the LLMs though... once you've made a habit of asking an LLM what it thinks about every little thing, it will feel more and more difficult to think for yourself. Absolutely do not use an LLM to do the writing practices I covered earlier, as that completely defeats their purpose.

Research iteratively

While you should do *some* research before making your first prediction on a question, don't feel like your research needs to be complete. Especially for topics that are evolving quickly, an important part of the research process is actually observing the way things *change over time*, and checking how well your early assumptions hold up as new information comes in. Use your initial research to learn enough about the topic that you can (a) make a reasonable first prediction, and (b) have enough background knowledge that you're able to parse new information as it becomes available and assess how it impacts your forecast.

Having ‘skin in the game’ can also be a useful way to prompt new lines of thinking on a question. I find that once I’ve made a prediction, it becomes easier to see the ways that prediction might go wrong. Making your first prediction, sitting with it for a day, and then re-assessing the next day is an effective way to see which areas you may want to research further.

Updating

Your forecast isn't final until the moment a question resolves. On prediction polling sites, you can usually update your forecast at any time and as many times as you'd like. In a prediction market, so long as you can find a trading partner, you can buy or sell shares until the market closes. So while it's good to come up with a reasonable first prediction, updating your forecasts appropriately in response to new information is just as important.

Updating is not strictly necessary in a prediction market. So long as you're making good bets (that is, when you buy shares, their expected value is greater than the price you paid), you'll make money in the long run. You would, however, expect to make *more* money by selling shares if the price corrects or buying shares in the opposite outcome if your original side becomes overpriced.

In a prediction poll, updating is a requirement if you want a good score. Your score is based on the relative correctness of your prediction compared to the community, *averaged over the duration of the question*. In the vast majority of cases, additional information becomes available throughout the question period that sheds new light on the eventual outcome. Even if your initial prediction is great, given the information available when you make it, it will become obsolete as new information arrives. If you don't incorporate this new information, your predictions will get steadily worse relative to the community.

You should update your forecast whenever your belief about the outcome changes. This could be when new reporting emerges, additional data is published, you do more research, or as time passes without anything happening. It could also be because you simply change your mind after thinking hard about the question or upon revisiting the question after a break.

To determine how much you should update your forecast in light of new evidence, consider both *how strong your prior belief was* and *how strong the evidence is*. If your prior belief was strong, and the evidence is weak, you shouldn't update much. If the evidence is strong, you should consider a larger update. When your prior belief was weak (maybe you couldn't find any good sources in your research or don't know much about the subject), you should be quicker to abandon your initial prediction in response to even relatively weak new evidence. This all sounds rather obvious, but you'd be surprised how helpful it is to simply ask yourself "how strongly do I believe this" and "how good is this evidence" and update your predictions accordingly.

Make note of what types of evidence would cause you to change your mind *ahead of time*, so that you're primed to respond if that evidence appears. People are great at finding creative ways to explain away evidence or contort it to fit their previously held beliefs, but it's harder to do that if we acknowledge ahead of time which types of evidence we'd find compelling.

Don't be embarrassed to make big updates, and don't be afraid to update frequently... your goal is to be correct, not consistent!



Example of forecast updating. Each dot is an update to my prediction. You can see how I reduced my forecast (gold dots) as time passed with no relevant news, and then how sharply I updated based on news that was released on December 25th.

Cognitive biases

A great deal has been written on an ever-expanding set of *cognitive biases* that human reasoning is susceptible to; the list is far too long to cover each one here. I'll narrow my focus to a handful of biases I notice frequently in my own thinking and which are especially relevant to forecasting, and will discuss specific strategies to manage each one.

Motivated reasoning

Motivated reasoning is when we've (consciously or not) decided ahead of time what conclusion we're going to reach, and we interpret any and all new evidence we encounter in a way that leads us to that conclusion. We downplay evidence that doesn't support our beliefs, and pay extra attention to evidence that does. We contort evidence that ought to *weaken* our belief into an argument that *strengthens* it. We seek out the best evidence that supports the conclusion we want to reach, but don't do the same for the other possible conclusions.

We're especially susceptible to motivated reasoning when we'd strongly prefer for one conclusion to be true, or when it would be inconvenient to change our mind. In forecasting, this can manifest as predicting higher odds of an outcome *we want to occur* or one that *we've staked reputation or credibility on*. In cases like these, we are highly 'motivated' to reach a particular conclusion, regardless of what the evidence actually suggests.

Time to talk about my worst run of forecasting to date. I made quite a few *terrible* predictions about the 2024 presidential election. I wanted Harris to win, predicted confidently that Harris

would win, and was proven badly mistaken. Looking back, I see a number of ways that I let motivated reasoning skew my reasoning:

- I placed extra trust in polls that showed Harris doing especially well, and dismissed polls that showed Trump doing especially well
- I convinced myself that polling error was more likely to be underestimating Democrats than Republicans
- I actively sought out analysis that claimed to show Democrats doing well in early voting, or which spoke favorably about their get-out-the-vote operations
- I watched the DNC, but not the RNC
- I downplayed the possibility that surges in Harris's polling numbers were due to temporary enthusiasm / response bias around the candidate switch

The issue here was not that I thought Harris was a good bet - I think there's a decent argument that prediction markets were overconfident in a Trump win, given the available evidence.⁵ Rather, the issue was with my reasoning *process* and with the *overconfidence* that process led to. As the points above show, I was not interpreting the evidence as a neutral observer, I was interpreting it as a Democratic voter hoping my side would win, and determined to find evidence to support that conclusion.

Avoiding motivated reasoning is a constant battle and there's not a simple 'fix,' but I'll offer a few strategies that can help you manage it in your forecasting:

- Rely as much as possible on base rates and other statistical evidence and apply intuition sparingly. Also keep in mind that motivated

⁵ Prediction markets favored Trump 65% to 35%, while election models like Silver Bulletin and FiveThirtyEight showed the race as a pure 50/50 tossup, and it was never clear why the markets were so bullish on Trump.

reasoning might lead you to look harder for the statistical evidence that supports your argument, so even this is not necessarily unbiased.

- Acknowledge which outcome you'd prefer (and how strong this preference is) and write this down when you make the prediction. Over time you can track if you're systematically overconfident in outcomes you want, and if you notice a trend, you can try and adjust for this up front
- Practice spotting motivated reasoning patterns in your forecasting or in your everyday thinking. Note when you're being dismissive of evidence that contradicts your beliefs, seeking out data that supports what you already think (without also looking for data that doesn't), or giving too much weight to evidence that someone with opposing beliefs wouldn't find compelling.
- Practice *steelmanning* different viewpoints. Often we weigh the *best possible argument* in favor of our preferred outcome against the *worst possible version* (the 'strawman') of the opposing argument. Instead, think through what the most charitable version of the opposing argument (the 'steelman') looks like and only judge your side against that.

Anchoring

'Anchoring' describes our tendency to stick close to the first number we see when trying to estimate a quantity or probability. Say you and several friends have a jar of jellybeans and you're trying to estimate how many beans there are in the jar. The first number that pops into your head is 5000. Before you say this out loud, your friend guesses 500. Now it's your turn to guess. What are you more likely to do, stick with your original guess of 5000, or guess something much closer to 500?

Most people are going to change their guess to be closer to 500, *even if they felt good about their original guess*. We feel psychological discomfort when we deviate significantly from the ‘public wisdom’, even when we have no idea if that public wisdom is well grounded

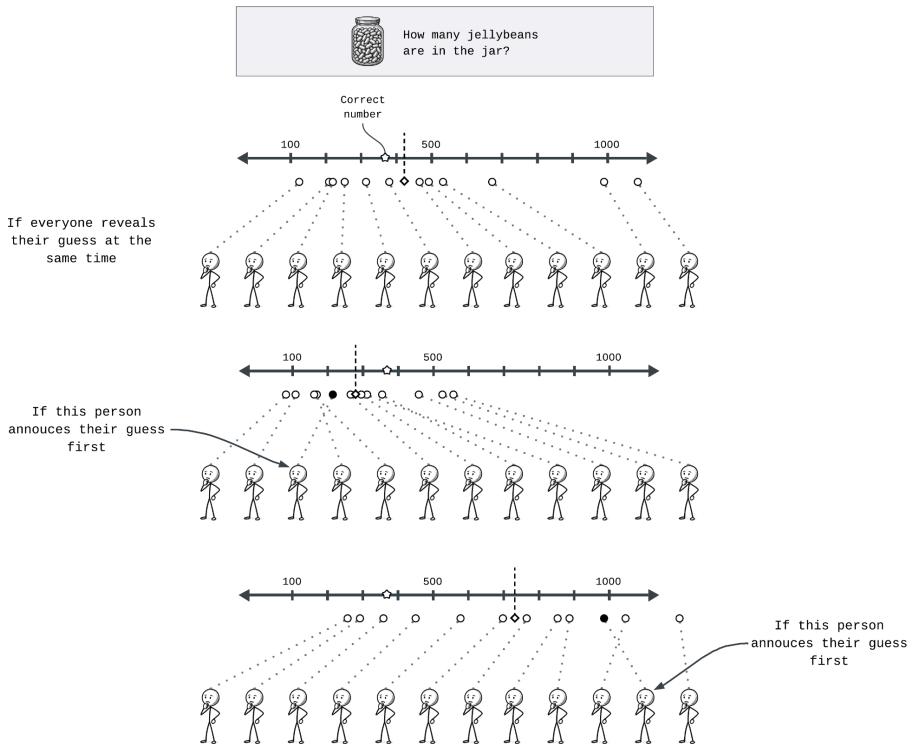
There are two main ways that anchoring comes up in forecasting. First, once we’ve observed the community average forecast (or the prediction market price), it’s much harder for us to make a prediction that significantly deviates from that number. Second, once we’ve made our first prediction on a question, it’s hard to update to a number far away from that original prediction. We feel an urge to ‘anchor’ to the first predictions we make or see, even when we encounter new evidence that invalidates our initial forecast or which should otherwise prompt us to significantly change our prediction.

The effect of this is that we react too slowly to changing evidence or make a prediction close to the community average, despite possessing a key insight that the community isn’t seeing. Worse, if *most people in the community are anchoring*, we can end up with severely distorted forecasts, even at the community level.

My best advice for avoiding anchoring is, on platforms where it’s possible, *keep the community forecast hidden by default* until after you’ve made your first prediction on a question. To be clear, you absolutely should look at the community forecast eventually. Often the community has picked up on important details that you missed, and in general it outperforms most forecasters on any given question. But you can’t unsee the community forecast once you’ve seen it. It’s useful to know what you think about a question personally, before you factor in public opinion.

When you aren’t able to hide the public forecast (you won’t be able to on prediction markets), or once you’ve already made

your first prediction, practice thinking through how much you want to change your forecast in light of new evidence *before checking how much the community reacted* to that evidence. Again, it helps to know what your personal intuition tells you, absent the influence of everyone else.



What ‘anchoring’ looks like in the jellybean example from earlier. Participants are influenced by the first person who announces their guess. If the first guess is especially low or high, this can skew everyone else’s guesses in that direction, since people tend to avoid guesses that are very different from the crowd.

Heads I was smart, tails I was unlucky

It's tempting to attribute our successes to skill and our failures to luck. When we predict something with confidence and then it happens, it's because of our incredible reasoning abilities. When it doesn't happen, it's because of that unlucky edge case that we of course knew was a possibility... our forecast was good, but sometimes the 5% chances happen, right?

The problem with this way of thinking is that it prevents us from learning from our successes and failures. If we misremember why we made the predictions we made, we'll misunderstand why they were right or wrong, and miss chances to improve the way we analyze future questions.

One way to mitigate this issue is to *write out why we're making the predictions we make* at the time we make them, and revisit those explanations once we know the outcome. After we know what really happened, we should make note of the *actual* reasons this outcome occurred, and check them against the notes we took beforehand. This helps us catch cases where our reasoning was poor, but we got lucky and guessed the correct outcome anyway, and also cases where our reasoning was good, but we truly were just unlucky.

Nothing ever happens

Because crowd prediction platforms want their questions to resolve within reasonable timeframes, they often apply arbitrary end dates, such as the end of a month or year. For example, you may see the question "When will the Russia-Ukraine war end?" posed as a series of Yes/No questions in the form of "Will the Russia-Ukraine war end by [date]?"

Prediction platforms also ask questions that are likely to capture attention, covering breaking news events and other topics circulating in popular discourse. Many of these markets speculate on the prospects or rare or outlandish things happening, things that would be major news stories should they occur.

Collectively, these two practices lead to an effect where most questions that are posed as “Will [X] happen by [date]?” ultimately resolve to No. This makes sense - a specific event can *fail to happen* repeatedly, but it can only *actually happen* once. The Russia-Ukraine war can fail to end every single month, but can only actually end once.

At the same time, people exhibit a strong *availability bias*, a tendency to overestimate the significance or likelihood of things they can easily imagine, or which are fresh in their mind. News coverage exacerbates this by specifically highlighting things that are unusual, that trigger strong emotional reactions, or are otherwise ‘newsworthy’.

The combination of these factors, the tendency of forecasting platforms to pose questions that usually resolve No and the tendency of news-readers to disproportionately notice rare and surprising events, has led to a popular rule of thumb among successful forecasters: “Nothing ever happens.” A surprisingly effective forecasting strategy is to find questions that are posed as “Will [X] happen by [date]?,” assume that the public is overestimating the chances, and predict a lower probability than what the public currently believes.

Will OpenAI announce an IPO before 2027? *Probably not.* Will Elon Musk form a new political party in 2025? *Probably not.* Will China invade Taiwan before 2028? *Probably not.* Will AI cause mass unemployment by 2030? *Probably not.*

The point I want to make here isn't that things actually never happen - they of course *do happen* all the time. Rather, it's that our biases generally push us in the direction of overestimating the chances of things happening, rather than underestimating those chances. It's easier to imagine rare things happening than to imagine them not happening, even though by definition, those things are rare!

A good way to mitigate the effect of availability bias, without just blindly predicting "No" on everything, is to lean heavily on base rates when possible. If you find a reliable base rate, consider that a strong prior, and only deviate significantly from it if the evidence telling you to is strong.

You should also consider whether the end date for a question has some special significance (perhaps there's an election that month/year, a legal deadline, a short time window for someone to act), or whether it's just an arbitrary end date, chosen to make the questions resolve within a reasonable time frame. Questions where the end date is arbitrary are especially susceptible to the 'nothing ever happens' effect, because there's no particular reason why any extra probability should be concentrated in the period leading up to that end date.

How to not lose all your money

In this final section, I'll talk about betting strategy for prediction markets, addressing the question of how much money you should spend if you think a market is mispriced. First though, I want to give a word of caution about prediction markets.

Prediction markets are zero-sum.⁶ Nobody makes money unless somebody else loses an equal amount of money. As of April 2025, only around 10% of traders on Polymarket were profitable - the remaining 90% had lost more money than they earned.

My honest advice to most people is not to trade in prediction markets at all, at least not unless you've established a track record of strong forecasting performance on a free platform like Metaculus or Manifold Markets. Or better yet, make 'fake trades' for a while to see if you could plausibly earn money in a real prediction market. Start with a bankroll of fake dollars and use a spreadsheet to 'buy' and 'sell' shares from the real order books for different questions on Kalshi or Polymarket. If you can't make a profit this way, there's no reason to think you will with real money.

I do think that by taking a principled approach, by practicing the forecasting strategies described in this guide, putting real thought and research into each question, updating your forecasts diligently, and making appropriately-sized bets, it's possible to earn a profit on prediction markets. But doing all of this takes effort, restraint, introspection, and patience. You shouldn't expect to just sit down, start firing off trades, and achieve any result other than losing all your money.

⁶ In fact Kalshi is actually *negative*-sum for traders because the platform charges a small fee on most trades (this is how the platform makes money). Polymarket has no transaction fees as of this writing, but presumably will add them in the future.

If you trade in prediction markets, you need to be honest with yourself about how skilled you are. Study your betting history and see if you have a real edge, figure out what topics you're good at and bad at, and never make excuses when you lose a bet. Assume that your counterparty in any trade is another intelligent human being who thinks that the price you're offering is advantageous to them, and make sure you have a sense for why they think that and why you think they're wrong. Give yourself hard limits and be prepared to quit if you lose more than a pre-determined amount of money.

And as I covered earlier in the guide, keep in mind that the benefits to developing forecasting skill are broader than 'making money on prediction markets'. It will help you learn about the world and make sense of current events, understand and overcome the biases that affect your reasoning, and make better decisions about the future. Even if your only goal is to make money from your forecasting talent, there are almost certainly easier ways for you to do that than by trading in prediction markets.

With all that in mind, let's move on to the problem of *bet sizing*.

Bet sizing

Say that a prediction market is trading at \$0.40 per share, but we are absolutely certain that the true probability of that outcome is 45%. In other words, the expected value of each share is \$0.45 and we can buy them for \$0.40. If we have a \$1,000 bankroll, how much of that bankroll should we spend ('bet') buying shares of this outcome?

One option is to *bet all of our money*. This runs the risk of us losing and going broke (there's a 55% chance of that), but it also

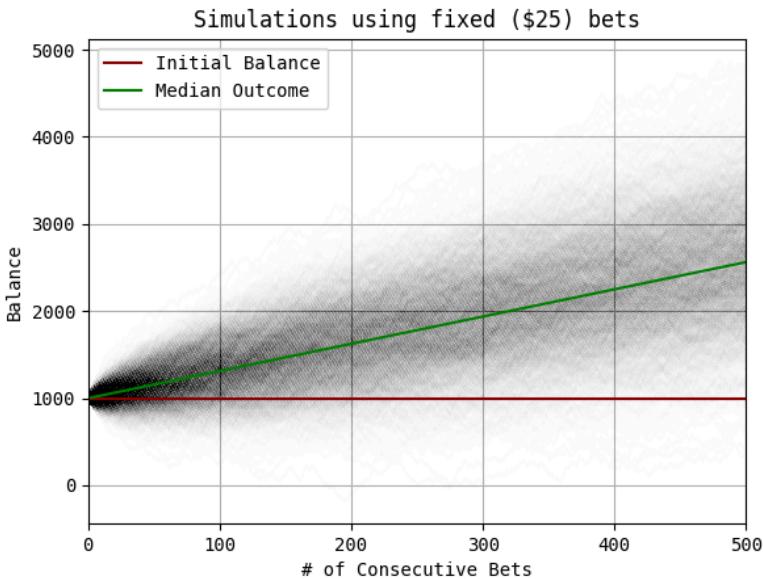
maximizes our expected profit from the trade. For every additional share we purchase, our expected profit increases by \$0.05, so shouldn't we just buy as many shares as possible?

If this is the only market we ever plan to trade in, and *if* our goal is truly to maximize the expected return *on this one single bet*, then yes, we should spend our entire bankroll on this market. However, if we plan to make more bets in the future, and we expect to have similar advantages in other markets, what we really care about is *maximizing the growth of our bankroll across many future bets*. Going all-in maximizes our expected profit on any single bet, but guarantees that we'll go broke if we apply the strategy repeatedly.

So we shouldn't bet our entire bankroll. What if we use a fixed bet size instead, say \$25? We know there's a 45% chance of us winning any given bet, and as long as we're able to find a similar edge on future bets, we'd have to lose *at least 40 times in a row* to run out of money, which is extremely unlikely.

(See the figure on the following page)

This strategy looks pretty good, doesn't it? Our bankroll grows steadily, and we've all but eliminated our chances of going bust! There's still a problem though... It's great that we're getting steady growth, but that growth is *linear*. Even a standard savings account offers us an annual rate of return, i.e. *exponential* growth. Our fixed-size bets will give us decent results early on, but over time it will get harder and harder to keep up with even modest compounding growth.



This figure shows the results of 2000 simulations in which we apply the ‘always bet \$25’ strategy. Darker regions indicate that more of the trials reached that region. Each simulation starts with an initial balance of \$1000, and makes 500 bets, wagering \$25 each time. For each bet, the ‘market price’ is \$0.40, but the true probability of winning the bet is 45%. After 500 rounds of betting, 1.25% of the simulations lost money, and the median final bankroll was \$2500

Clearly there must be a better way to get value out of our advantage. Intuitively, a good betting strategy should have the following characteristics:

- We bet more money when our edge (the difference between our forecast and the market price) is larger
- The size of our bets scales with our bankroll. If our bankroll grows, we should place larger bets, and if it shrinks we should place smaller bets

- Our bet sizes are small enough that a string of bad luck won’t eliminate all of our gains

In 1956, John Larry Kelly Jr. described a strategy, now known as the *Kelly Criterion*, which has all these characteristics.

The Kelly Criterion

Kelly’s formula tells us *the fraction of our bankroll to wager* on a bet in order to maximize the long term expected growth rate⁷ of that bankroll. The formula is typically presented as:

$$f^* = p - \frac{1-p}{b}$$

Where f^* is the fraction of our bankroll to bet, p is our ‘true probability’ of winning (our forecast), and b is the proportion of a bet that we receive if we win. For a bet where we’re offered 3-to-1 odds, $b = 3$, we receive \$30 plus our original \$10 if we bet \$10 and win. In the context of prediction markets, where the ‘odds’ are presented as a probability, it’s simpler to represent b in terms of the prediction market price, which yields this formula:

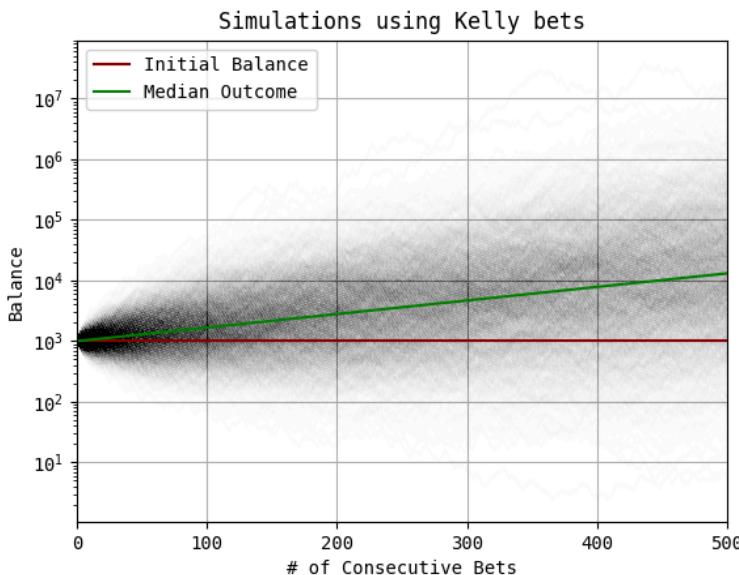
$$f^* = \frac{(p_{forecast} - p_{market})}{(1 - p_{market})}$$

Let’s apply this to our earlier example. p_{market} is 40% and $p_{forecast}$ is 45%, which gives us:

$$f^* = \frac{(0.45 - 0.4)}{(1 - 0.45)} = 8.0\%$$

⁷ ‘long term’ = ‘across a large number of bets’, ‘expected’ = ‘if lots of people followed this strategy, what would the average result be’, ‘growth rate’ = ‘the percentage that our bankroll grows by with each bet’. So ‘long term expected growth rate’ is the average percentage our bankroll would grow by on each bet, if we take a large number of bets.

So the Kelly Criterion tells us to spend 8% of our bankroll on this bet. We use this fraction regardless of the size of our bankroll. If we have a string of wins, the *fraction* we should bet will stay constant, but the *size* of the bet in dollars will increase. Likewise, if we lose a bet, the dollar amount we wager the next time will be smaller. Let's see how our simulation looks when we use 'Kelly bets':



Results of 2000 simulations where we use the Kelly Criterion to determine our bet sizes. Like before, we know the true chance of winning is 45%, but the market price is \$0.40. Note the log scale on the y-axis. After 500 bets, the median final balance is \$13,000, and 13.1% of the simulations ended with less money than they started with.

Now *those* are some excellent results! After 500 bets, the median outcome is that our initial bankroll of \$1,000 has grown to \$13,000, and 25% of the simulations are north of \$67,000 (one lucky simulation crossed \$17,000,000). All that from being just 5% more confident in the correct outcome than the market.

	Forecast Probability																			
	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100%
5%	0%	5%	11%	16%	21%	26%	32%	37%	42%	47%	53%	58%	63%	68%	74%	79%	84%	89%	95%	100%
10%		0%	6%	11%	17%	22%	28%	33%	39%	44%	50%	56%	61%	67%	72%	78%	83%	89%	94%	100%
15%			0%	6%	12%	18%	24%	29%	35%	41%	47%	53%	59%	65%	71%	76%	82%	88%	94%	100%
20%				0%	6%	13%	19%	25%	31%	38%	44%	50%	56%	63%	69%	75%	81%	88%	94%	100%
25%					0%	7%	13%	20%	27%	33%	40%	47%	53%	60%	67%	73%	80%	87%	93%	100%
30%						0%	7%	14%	21%	29%	36%	43%	50%	57%	64%	71%	79%	86%	93%	100%
35%							0%	8%	15%	23%	31%	38%	46%	54%	62%	69%	77%	85%	92%	100%
40%								0%	8%	17%	25%	33%	42%	50%	58%	67%	75%	83%	92%	100%
45%									0%	9%	18%	27%	36%	45%	55%	64%	73%	82%	91%	100%
50%										0%	18%	20%	30%	40%	50%	60%	70%	80%	90%	100%
55%											0%	11%	22%	33%	44%	56%	67%	78%	89%	100%
60%												0%	13%	25%	38%	50%	63%	75%	88%	100%
65%													0%	14%	29%	43%	57%	71%	86%	100%
70%														0%	17%	33%	50%	67%	83%	100%
75%															0%	20%	40%	60%	80%	100%
80%																0%	25%	50%	75%	100%
85%																	0%	33%	67%	100%
90%																		0%	50%	100%
95%																			0%	100%

fraction of bankroll
 to wager, according to
 Kelly Criterion
 

This table shows the Kelly fraction (the percentage of one's bankroll to wager on a bet) for different forecast probability / market probability combinations. Betting the Kelly fraction gives the optimal long term expected growth rate. For the reasons discussed below, full sized Kelly bets should not typically be used in practice.

Notice that when we use the Kelly Criterion formula, as our forecast gets farther from the market price (as our edge gets larger), the fraction recommended by the formula increases. The numerator ($p_{forecast} - p_{market}$) grows, while the denominator ($1 - p_{market}$) stays the same. This aligns with what our intuition told us about good betting strategy, that we should bet more money when our edge is larger.

The problem with going ‘full Kelly’

Unfortunately, there’s an assumption we’ve been making up to this point which will cause major problems for us if we actually use *full Kelly* bets, that is to say we bet the full percentage of our bankroll recommended by the Kelly Criterion. *We’ve been assuming that our forecasts are perfect.*⁸

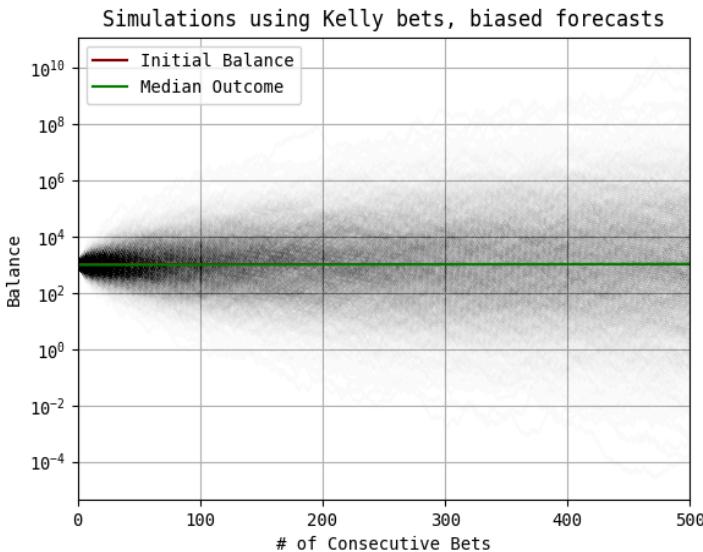
The Kelly Criterion tells us how to maximize the growth of our bankroll, if we know with 100% certainty that our forecast is exactly correct. This is never the case in reality, and if we bet as though it is, the results will be disappointing.

Returning to our example, let’s say that while the market price is \$0.40 and the ‘true probability’ is still 45%, our forecast says 50%. Our forecast is directionally correct (we’ve spotted a market that really is underpriced) but we’ve overestimated the true odds. What happens if we assume our 50% forecast is perfect, and make full Kelly bets accordingly?

Even though in some sense we’re making ‘good’ bets – each individual bet still has a positive expected value, and we’ve correctly recognized that the market price is too low – our forecast error means that our losses now punish us just as much as our wins reward us. The average growth rate across these simulations is 0, we are equally as likely to lose money as we are to gain it.

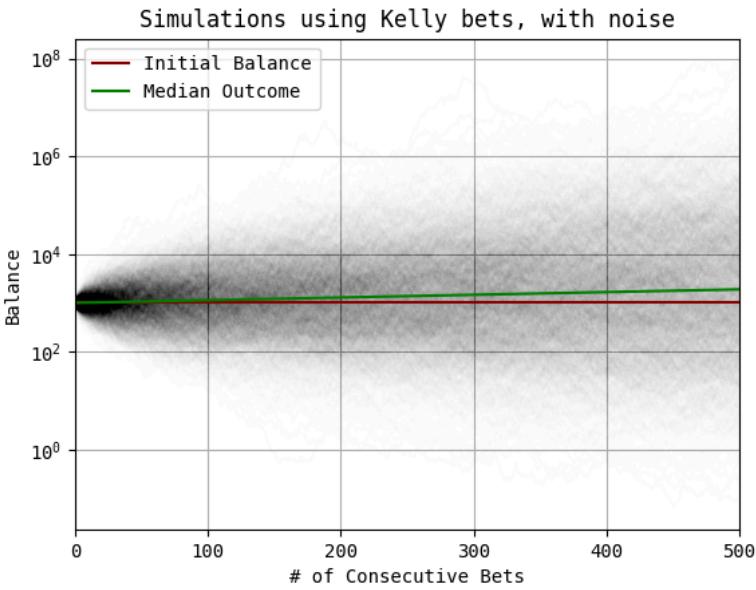
If our forecast is just slightly *more* incorrect than this, say it says 51% instead of 50%, or the true probability is 44% rather than 45%, our expected growth rate becomes *negative*. If we’re *very* wrong, that growth rate becomes *very* negative.

⁸ We’ve also been assuming that we’ll keep getting offered bets, and be able to increase our bet sizes indefinitely, even while maintaining a significant edge. If we really were to start winning larger and larger sums of money, our counterparty would catch on sooner or later and stop offering us these bets.



Results of 2000 simulations where the market price is \$0.40 and the true probability is 45%, but we size our Kelly bets as though the probability is 50%. We're correct that the market price is off, but we've overestimated the true odds. We now lose money in 48% of trials, and our median bankroll after 500 bets is just \$1070. Our forecast error has negated any advantage we had.

But maybe this isn't such a big deal... as long as we don't *systematically* overestimate the true odds, shouldn't we be okay? Say that sometimes our forecast is a little higher than the true odds and sometimes it's a little lower, but in roughly equal amounts, and on average our forecasts are spot on. Returning to the previous example (\$0.40 market price, 45% true probability), say our forecast correctly gives 45% on average, but follows a distribution with a standard deviation of 5%. There's no systematic bias in our forecasts, and about 70% of our predictions are within 5% of the true probability. As the following simulation shows, even this modest amount of 'noise' in our predictions makes it much harder to grow our bankroll.



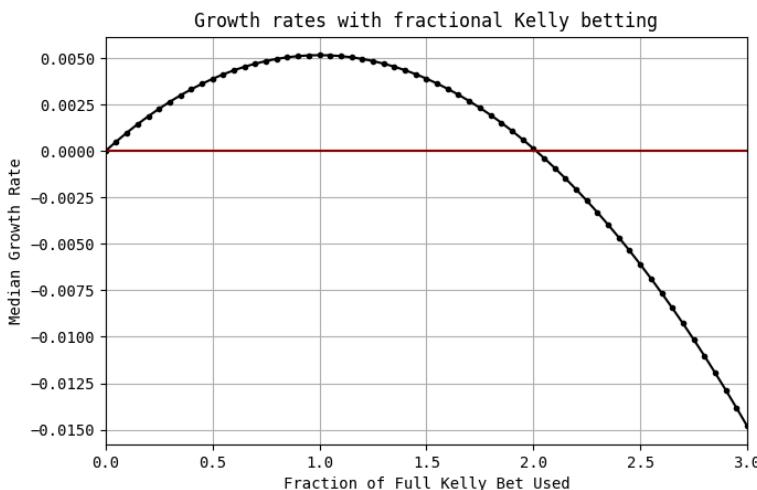
Results of 2000 simulations where the market price is \$0.40, the true probability is 45%, and our forecasts are correct on average, but noisy. For each bet, our forecast probability (which determines our bet size, using the Kelly Criterion) is drawn from a normal distribution, centered at 45%, with a standard deviation of 5%. Adding this noise, we now lose money on 42% of trials (compared to 13.1% without the noise), and our median final bankroll is \$1,900 (compared to \$13,000 without the noise).

Now it shouldn't be surprising that adding noise to our forecasts reduces our winnings. The Kelly Criterion gives us the best possible growth rate, so if we're sometimes betting above or below the fraction it recommends, we would expect to do worse than if we use the optimal fraction every time. So we should probably temper our expectations a little. But if we dig a little deeper, we can find a strategy that still manages to do quite a bit better than this last example.

It turns out that most of the damage is being done by our very worst forecasts, the ones where we're overestimating the odds the

most and betting way too aggressively as a result. If bets exceed the Kelly fraction by too much, they start to *lose* us money on average, even though we've picked the correct 'side' of the bet. The farther beyond the Kelly fraction our bet sizes go, the faster we lose money. These losses can easily get so large that they cancel out all the winnings from the rest of our bets.

As the following chart shows, bets that are more than 2x the Kelly fraction will lose us money in the long run, and the penalty for betting too much increases rapidly beyond that point:



This figure shows the expected per-bet growth rate of a bankroll, using different 'fractional' Kelly bets. For each fraction from 0x to 3x the 'full' Kelly bet, we run 5000 simulations (each consisting of 500 bets with \$0.40 market price and a true probability of 45%) where every bet is scaled up or down by that fraction, and plot the median growth rate from those simulations

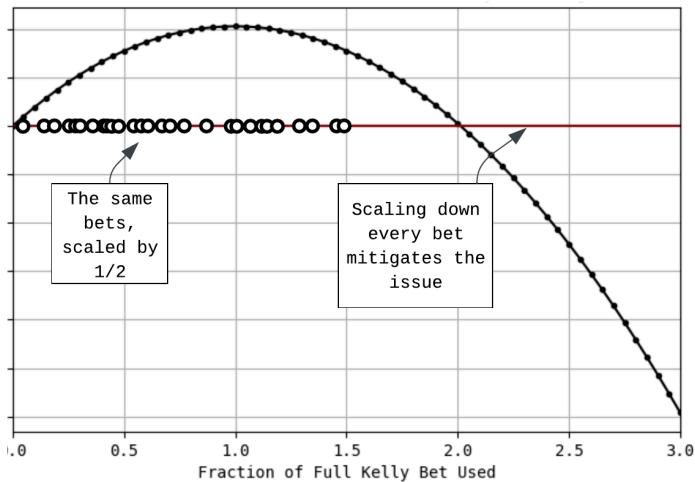
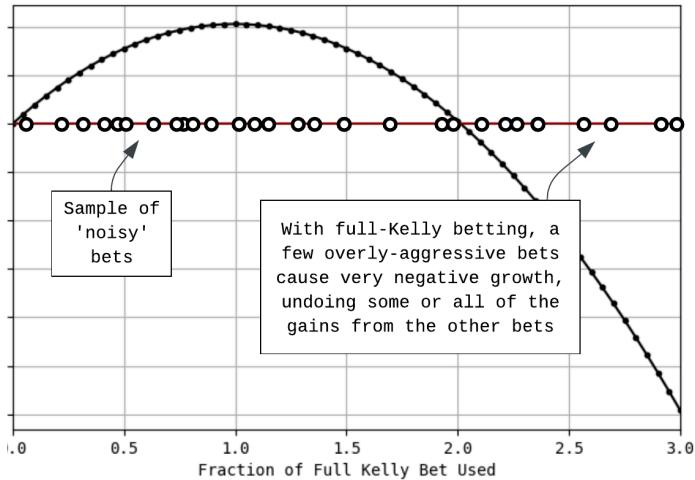
In our toy example, where the market price is \$0.40 and the true probability is 45%, a forecast of just 49% already lands us at 2x the 'correct' Kelly fraction, and a forecast of 52% puts us at 3x the proper bet size, losing us more money than an optimal bet would

earn us. It doesn't take much noise in our forecasts to start making losing bets.

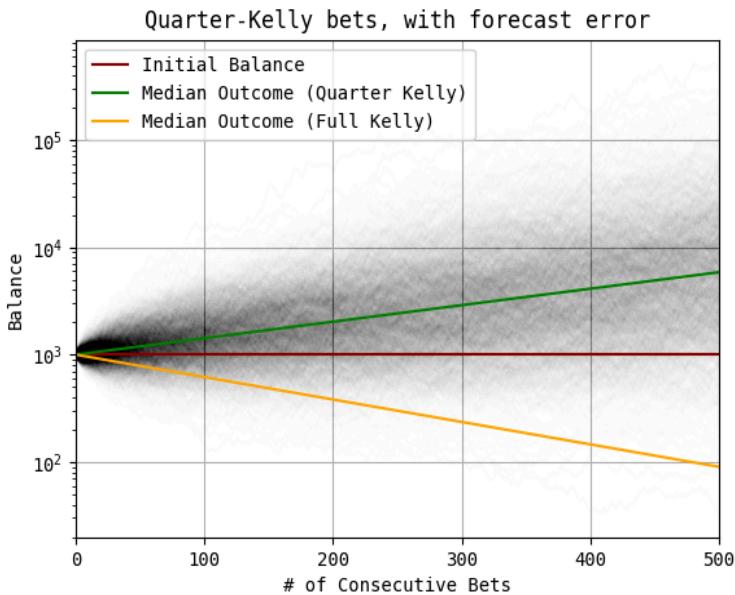
On the other hand, notice that no matter how far *under* the Kelly fraction our bets are, they still give a positive long term growth rate. If our bet sizes fall anywhere between 0 and 2x the Kelly fraction, we break even or turn a profit. While overly aggressive betting can result in very negative growth rates, the worst we can do by being conservative in our betting is break even. The growth rates achieved by betting half or even just a quarter of the optimal amount are still quite good.

For these reasons, we're likely to see better results if we make *fractional Kelly bets*, perhaps half-Kelly, quarter-Kelly, or even eighth-Kelly bets. Though we should work to make our forecasts more accurate over time, we'll never completely eliminate errors like bias and noise. Because *overestimating* our edge and betting too aggressively is more harmful than *underestimating* our edge, scaling down our bets helps to mitigate the damage from our worst bets, which otherwise might undo the gains from our good bets.

I'll emphasize that *all of this assumes you actually have an edge*. This should be obvious, but if you are picking the losing side of bets, it doesn't matter how you size your bets, you are going to lose money over time.



Fractional Kelly betting (betting $\frac{1}{2}$ or $\frac{1}{4}$ of the 'full' Kelly fraction recommended based on our forecast) can improve our results by mitigating the effects of random error in our predictions. Occasionally overestimating our edge by a large amount can cause overly-aggressive betting that loses money, cancelling out the gains made on other bets. Scaling each bet down reduces these losses, without eliminating the gains from the good bets



Results of 2000 simulations where the market price is \$0.40 and the true probability is 45%. In these simulations, we assume both bias (forecasts overestimate the true odds by 5% on average) and noise (forecasts are drawn from a normal distribution with a standard deviation of 5%). Because of this forecast error, full Kelly betting gives a median final balance of just \$90. If we use quarter-Kelly bets instead, our median final balance is \$6,300.

I've covered a lot of ground here, so I'll close by re-stating some of the key ideas from this section:

- Because we can reinvest any earnings, having a consistent edge should lead to exponential, rather than linear returns. Our goal is *compounding growth*.
- To achieve exponential returns, we need to bet *fractions of our bankroll* rather than fixed dollar amounts. The dollar size of our bets should scale up and down with our bankroll.
- If our bets are too large, we will lose money over time, *even if we pick the correct side on every market*.

- The Kelly Criterion tells us the optimal size for our bets, *assuming a perfect forecast*. Betting far beyond the Kelly fraction causes us to lose money, while the worst we can do by betting less than the Kelly fraction is break even.
- Our forecasts are not 100% perfect. Sometimes we'll overestimate our edge and sometimes we'll underestimate it. Overestimating our edge can lead to bets that are so aggressive, they lose us money, perhaps more than we earn back from our good bets.
- Half-Kelly or quarter-Kelly bets reduce our chances of making these overly-aggressive bets, and are likely to achieve better results in the long run than full Kelly betting.
- **No betting strategy will work if you don't have an edge over the market.** You should record every bet you make and determine whether you are correctly spotting mispriced markets. If you aren't, you should stop betting.

Recommendations for Further Reading

Superforecasting

Dan Gardner and Philip E. Tetlock

The Signal and the Noise

Nate Silver

Everything Is Predictable

Tom Chivers

Guesstimation

Lawrence Weinstein and John Adam

The Scout Mindset

Julia Galef

Writing to Learn

William Zinsser

Fooled by Randomness

Nassim Taleb