

RS 145: Phil Tetlock on, "Superforecasting: The Art and Science of Prediction"

Julia: Welcome to Rationally Speaking, the podcast where we explore the borderlands between reason and nonsense. I'm your host, Julia Galef, and with me is today's guest, professor Philip Tetlock.

He is a professor at the University of Pennsylvania where he serves in the Departments of Psychology, Political Science, and then also the Wharton School of Business. He is the author of many books, including famously, *Expert Political Judgment: How Good Is It? How Can We Know?* And his most recent book, *Superforecasting: The Art and Science of Prediction*.

Philip is also the co-founder and co-principal investigator of the Good Judgment Project, which is a long-term study on how to improve the accuracy of human judgment and forecasting. Welcome to Rationally Speaking.

Phillip: Thank you.

Julia: We're going to be talking today about Superforecasting and what you've learned from running the Good Judgment Project. Probably a good place to start is to say that many of our listeners, I'm sure a large chunk of our audience, is already familiar with you and your work to at least some extent. Probably many of them associate you most strongly with the finding that received quite a lot of publicity in the years since *Expert Political Judgment* came out -- that experts, forecasters of economic or political topics, are not really experts. That their judgments, their forecasting is no better than -- well, sometimes the claim goes no better than amateur, sometimes it goes no better than chance.

In the beginning of your new book, Superforecasting, you talked about how that finding has sort of gotten distorted from the actual conclusion that you found in your research. Why don't you kick things off by elaborating a little bit on that?

Phillip: Sure. I think I started the mischief by using the "dart throwing chimpanzee" metaphor in the book, although I don't think the metaphor was unique to me. I think the Wall Street Journal and The Economist have used it in picking on stock pickers.

Julia: What is that metaphor? Just to state it.

Phillip: The dart throwing chimpanzee. We didn't actually have a dart throwing chimpanzee, it's just shorthand talking about how well you would do if you were generating predictions purely by chance.

Julia: The metaphor, then, is that the experts were no better than dart throwing chimpanzees?

Phillip: Well, that is the takeaway line a lot of people have taken from the Expert Political Judgment book. I can understand why some people have taken it that way. An

important subgroup of experts did indeed have a lot of difficulty performing appreciably better than chance. That is true. But some experts did perform better than chance.

Experts can perform better on some tasks than on others. Just as it's easier to see an eye chart if you're close up than if you're far away, it's easier to see into the future one year out than five years out. Whether experts outperform chance or not hinges on how far out in time we're asking the experts to see. It also depends on which subgroup of experts we're talking about in the original Expert Political Judgment sample.

Julia: Your work on the Good Judgment Project was in some sense a follow-up to your work that went into Expert Political Judgment -- do you want to explain the question that you were investigating in the Good Judgment Project?

Phillip: Sure. One way of characterizing the earlier work and the later work is that the Expert Political Judgment work, the early work, was more about cursing the darkness. And the later work, the Good Judgment Project, is about lighting candles.

Julia: That's a great way to put it.

Phillip: It was a research program that originated in a big tournament sponsored by the Research and Development Branch of the Office of the Director of National Intelligence, IARPA, Intelligence Advanced Research Projects Activity.

It was a quite extraordinary thing for a government bureaucracy to do when you think about it. I think it had its roots probably in the major intelligence failure in 2003 on weapons of mass destruction in Iraq. I think that event and some other events increased interest within the intelligence community in thinking more systematically about how they use probability judgments.

The IARPA tournament, in a sense, was an effort to explore the degree to which it's possible to make meaningful probability judgments of geopolitical problems, economic problems, military problems that are of direct national security relevance. They posed approximately 500 questions over four years and the forecasters we recruited, tens of thousands of forecasters actually, attempted to answer those questions.

The competition itself involved five different academic teams at the beginning. The Good Judgment Project was one of four university-based teams competing in the tournament. The Good Judgment Project happened to win. The book, in part, is about how it won that forecasting tournament.

Julia: You won by a pretty significant margin, didn't you?

Phillip: Yes, we won by such a significant margin that everyone was surprised. We were certainly among those who were very surprised by the size of the margin. IARPA thought that it would be ... IARPA set the performance goals at the beginning. They said, "Look, if you guys can beat the average forecast from a control group of forecasters by 20% in year one, 30% in year two ... I think it was 40% and 50% in years three and four,

something like that. If you can beat the “wisdom of the crowd”... by those margins we'll be really happy and that would be great. We'll keep funding all the teams and so forth.

Our best forecasters, and best algorithms, managed to beat the 50% benchmark after year one. They were the only team to beat benchmarks consistently in both years one and two.

IARPA decided to consolidate resources. We could transfer some of the people who were on the other teams into our team, and we would then compete as a mega-team against a prediction market and some benchmarks inside the intelligence community itself.

Julia: Let's just back up briefly to get a little more background. Can you tell us who was the control group that you beat? Who were the forecasters that you outperformed? Also -- can you give a couple of examples of the kinds of questions that your forecasters were answering?

Phillip: Oh sure, sure. We asked early on whether Medvedev and Putin would change jobs and Putin would get his old job back. We asked about whether Assad would fall. We asked whether Greece would leave the Euro-zone.

All these questions had to be asked in a very specific way. Would this very specific thing happen by this very specific date. That's crucial, because often in real-life forecasting people have too much wiggle room to argue that they were right after the fact. We wanted questions that were so well-defined that there would be virtually no wiggle room for reasonable people to disagree about what happened after the fact.

We called that the clairvoyance test. That was a very important part of the forecasting tournament, the generation of questions that are rigorously resolvable, and at the same time relevant to national security.

Julia: Yeah, just on that point, I really appreciated that part of the book where you talk about the importance of being concrete and concise enough in your predictions that you could even tell if you got it wrong. Especially when you probably have some emotional investment in being able to conclude that you got it right after all.

I was talking to a friend of mine who is the CTO of a big tech company out here in the Bay, and I was asking him what he looks for in hiring people. And the biggest thing that he emphasized was he really hates when people won't, as he said, “stick their neck out” and be concrete or precise. I forget what word he used.

I double-clicked on that phrase, “stick your neck out” -- why did you use that phrase? He said "Well, the more concrete you are in what you're claiming, the greater chance people will be able to tell that you're wrong."

Phillip: Yes, exactly, and the greater the opportunity you'll have to learn how close or far you are from the truth, and to learn. If you keep your predictions vague it's very, very

difficult after the fact to figure out how close or far you were, and whether your beliefs need to be modified in one direction or another. You cut yourself off from really valuable feedback by relying on vague verbiages.

Julia: Oh, then the other part of my question was who was the control group that you beat?

Phillip: That was a group that was recruited by the government. The forecasters were recruited in essentially the same way that our forecasters were recruited, by advertising in professional societies and so forth.

Julia: Great. So the forecasters who you recruited were essentially amateurs? Probably more knowledgeable amateurs than the average person, but not people who were making a living out of forecasting?

Phillip: Well, they certainly weren't pundits. They certainly didn't make their living ... Very few of them made their living writing or commenting on contemporary political, economic, military trends.

There were a few who were former intelligence analysts and so forth. There were people who had some significant subject matter expertise. When you consider the range of questions that IARPA asked, it's virtually impossible for any human being to be considered a subject matter expert on more than a tiny fraction of them. Imagine going from a question on what's the Arctic sea ice mass going to be next fall, to a question about Syrian refugee flows, to a question about whether there's going to be a violent China-Japanese clash in the East China Sea, to a question about what's going to happen to the Chinese currency, and then move on to the Guatemalan elections.

Julia: Right, right.

Phillip: All over the map. Literally all over the map.

Julia: Actually, full disclosure: I know several people who were participating in The Good Judgment Project. And I do know that they spent a fair amount of time just keeping informed, and taking in new information that might be relevant to these predictions.

Phillip: I'm glad that you do. It's always fun to talk to the participants in The Good Judgment Project, not just the superforecasters. I mean there are a lot of extremely talented people who didn't make the cut for the superforecaster, which is somewhat arbitrary. We were really blessed by the variety of talented people who signed up and dedicated a lot of their time to thinking about these problems and the challenges of attaching meaningful probability estimates to real world events that many pundits believe it's impossible to attach probability estimates to.

Julia: Yeah. At some point later in the conversation I hope we'll get to talk about the motivation or the drive that animated the superforecasters. For now, I just want to make sure we get to the, I guess, the "punchline" of what you've found was common to

these superforecasters -- what it seemed that they were doing in order to be superforecasters.

Phillip: Well, if I had to identify one mega takeaway from the project it would be really pretty simple. It would be that if you want to become a superforecaster, a necessary, although not sufficient, condition for becoming a superforecaster is the following: You have to believe that it's possible to cultivate the skill of subjective probability estimation of massive real world events, and you have to believe that it's worth your effort to give it a try.

We have a quote in the book from a great poker player and also a chief risk officer of a big hedge fund, AQR, Aaron Brown, about how you tell the difference between a great poker player and a talented amateur.

The short answer is, in his view, that the great player knows the difference between a 55/45 proposition and a 45/55 proposition. Now when you're talking about poker, people say, "Well sure." Nobody really wants to argue with that because poker is a game in which you're sampling from a well-defined universe of cards, the card deck. It's a game of repeated play, over time, the superior acumen can translate into better outcomes, even though the better player might well lose on any given hand.

Is the same true for making predictions about Syria? Is the same true for making predictions about Russian geopolitical intent? Or whether Scotland is going to secede from the UK and so forth? Well, some other people say, "Well, history is different, it's not like sampling cards from a poker deck. History only runs once. We only get to see one run of history.

There are some very smart people who are very suspicious of the idea that it's even meaningful to attach probability estimates to the questions that were posed in the IARPA tournament. What I regard really as the most extraordinary result of the tournament is that we have shown pretty conclusively that the probability judgments our forecasters were making were meaningful.

Julia: Yeah, I agree that's very interesting. Both that you were able to demonstrate it and also that many very smart people don't have that intuition, which I've also encountered in the work that I've done trying to train good judgment and rationality.

The only way that I've been able to sometimes get traction on that -- when someone is insisting that there is no way to assign a number to their degree of belief, or to their prediction -- is to say, "Oh, okay, then it's probably about 95%? Then they'll recoil and say "No it's nothing like 95%!" And I'll say, "Okay, so is it like 2%?" They'll say, "Well, no, that's crazy. That's far too low." And I can gradually coax them towards at least a range where they won't say it's absurd. That's the only way to do it.

Phillip: That's a wonderful example, and it's related to something we're about to do on the Good Judgment blog. We're about to issue a challenge to some of the skeptics... to make a series of bets that the superforecasters can outperform chance on questions

that the skeptics themselves nominate as being particularly unique. Aaron said that uniqueness can't be a matter of degree, it has to be all or none, but I think one of the clear takeaways from my research is that uniqueness is very much a matter of degree.

Julia: That's right. This actually leads me to a question I was looking forward to asking you that I currently feel rather confused about. Which is that, despite my resolve that you can assign numbers to your degrees of uncertainty, and that it's useful to do so, it still seems to me that there is a pretty significant difference between the uncertainties that I have in some domains, and the uncertainty I have in other domains.

For example, the 50% that I would assign to a coin flip coming up heads feels very different than the 50% I would assign to ... just to pull up a made up example, human level artificial intelligence being invented in the next 100 years, something like that. Where maybe I would say 50% to that, but it feels like a very different kind of 50% than the 50% I'd give to the coin flip.

In some sense, mathematically, it's not very meaningful to talk about your uncertainty about your uncertainty. It just sort of cancels out, like layers of tetris blocks or something. I've been searching for a way to capture what feels to me like a real difference in those kinds of estimates.

The only thing I've come up with so far, which -- I'm curious what you think about this -- is that the difference is something like the probability I assign to my current estimate changing in response to new information in the future. In that I would be very surprised if the 50% that I put on the coin flip coming up heads would change in response to new information. But I wouldn't be surprised at all if my 50% on the artificial intelligence question would change. Of course I don't know which direction it would change, because if I did then it would already have changed. But I expect the magnitude of the change to be greater or something like that. What do you think?

Phillip: That's very interesting. A lot of philosophers have talked about these issues. We use some philosophical terms briefly in the end notes of the book, distinguishing between epistemic and aleatory uncertainty. I'll come back to that in a second. Your example of the coin flip vs the AI, the projections of the future of AI, yeah, those do feel different. I'm not sure of the difference is, what you're pointing to, the ...

Julia: Yeah, nor am I.

Phillip: Imagine changing your mind in response to new evidence. You could imagine a coin flip as being conducted by a demon that has the power to change at any given moment whether the coin is biased or not, but you're starting from a 50/50 prior. There you would be willing to change your mind, right, quite quickly. Would that render the coin flip problem interchangeable with the AI problem? I think so.

Julia: I mean it feels like it comes down to how well defined ... Maybe it's about how well defined the prior is or something?

Phillip: Yeah, I think so. Yes. How much ... Right, you say a lot of people are very dubious about second order probabilities but you might say "I think there is a 50/50 probability" with respect to the AI scenario, but what you really mean is "I think the probability could be anywhere between 10 and 90." The other is you think it's more like 50/50, so it would be a range of probabilities that you're endorsing rather than a scalar.

Julia: My best counter to my own claim that you can put a probability on anything is the fact that -- and I'm wondering if you have a counter to this counter -- if you really feel like you know nothing, then you kind of fall back on some sort of ignorance prior, which in the case of a coin flip would be 50/50, "I know nothing."

...Well, your ignorance prior will change depending on what kind of question it is. There are lots of different ignorance priors you could use. There are a lot of different reference classes that you could say the case of the development of artificial intelligence belongs to. Which reference class you pick would affect which ignorance prior you use. But it seems kind of like there is no bottom to that process where you can have a prior over different ignorance priors -- but where does that come from, etc., etc.? At a certain point I'm sometimes inclined to say that there are probabilities that aren't well defined at the bottom.

Phillip: Fair enough. Yes. I don't know whether you're readers are familiar with ... You're using the term ignorance prior, maybe you want to define it for your listeners or not but ... I said readers as if I'm writing a book.

Julia: Yeah, I make that mistake sometimes, too. I think it's because the content is so cerebral that it really feels like a book in audio form.

1 The ignorance prior is just the probability distribution that you have when you don't have any particular information about the situation aside from just what kind of situation it is.

Phillip: Yeah, or even more radically, you don't have any information at all aside from the fact that a question was asked. The ignorance prior in the IARPA questions is: is something going to happen, or not going to happen? As opposed to sort of a multinomial where you have many options. On the dichotomous IARPA questions, the ignorance prior is 50/50.

The ignorance prior is really mischievous because you can take the question, and if you carve it up into ... If you carve one side of it up into more compartments than the other side, you'd weight the probability more toward that side. Do you see what I'm saying? They did a famous demonstration on support theory by Craig Fox and Amos Tversky and people like that. It's very mischievous.

Here's a takeaway on ignorance priors, though. The worst forecasters in the IARPA tournament were much likelier than superforecasters to use 50/50. They are much likelier to use ignorance priors.

Why didn't the superforecasters retreat to the ignorance priors at least when they were initially asked the question? Because after all, even the superforecasters aren't going to know very much about many of these questions.

The answer to that seems to be the superforecasters are really quick to look for reference classes or comparison classes that help to narrow the uncertainty range and put them in a more plausible ballpark than 50/50. If you're asked the question, "Is some obscure African dictator is going to be in power after 12 months have expired," if you were a regular forecaster you might say, "I have no idea, 50/50."

If you are a superforecaster, you're more likely to say, "Well, let's see. How often do dictators in general who have been in power X years tend to survive another year," and the answer is probably about 90% plus. In which case the superforecaster would be more likely to start with a very high probability for longevity in power and then adjust that probability in response to developments. Like, well, it turns out that particular African dictator is 91 years old and has advanced prostate cancer, or that particular African dictator there's riots in the capital city. You see what I'm saying?

Julia: Right, right, right. That brings up another interesting question that might occur to people as they read your book, although you do address it in the middle of the book, which is ... I guess it has two parts. One is how confident can we be that the people who were identified as superforecasters weren't just, (a) lucky, because if you have thousands of people making predictions then even if none of them have special skill, just by chance some of them are going to end up making more accurate predictions than the rest. Or, (b), just the people who spent more time on it. Who actually took the trouble to investigate the base rates in that reference class or to look at other reference classes?

Which is still a valuable thing to learn if that's what is going on, but it doesn't really point to the hypothesis about there being an additional skill beyond just taking the time to investigate the questions.

Phillip: You're asking two questions there. One is a regression toward the mean and the other is about the cost benefit, use of the time and effort. The first question is in a sense is the Warren Buffet or George Soros question -- that, you toss enough coins thousands of times and some are going to wind up heads 60, 70, 100 times in a row. That's just going to happen. Then you're going to call those people geniuses because they happen to be in the right place at the right time.

Most investors don't subscribe to such a strong form of the efficient market hypothesis that they would dismiss the very best investors as just lucky. It's an interesting idea.

We tested it very systematically in the IARPA tournament by looking at how much regression toward the mean there was in performance one year to the next. If the superforecasters were performing on a task of coin toss guessing, and it's a fair coin... Let's say we take the top 2% of forecasters in the coin toss guessing in year one, and we say, "Well, what percentage of those are going to wind up in the top 2% in year two?"



The answer would be maybe 2%. The vast majority of the forecasters are going to regress toward the mean, right? The best prediction of where the superforecasters in year one are going to be in year two is the mean because there is zero correlation, right, between how well you do from one year to the next. It's just going to be total regression toward the mean.

Julia: I mean unless you're gaining skill as you do it, right?

Phillip: Well, unless you believe in ESP or unless the coin is somehow rigged, right?

Julia: Oh, with the coin, yes.

Phillip: In the coin case. If you look at it ... The question is what's the ratio of skill to luck in performance in the IARPA tournament. You say, what's the ratio of skill to luck in batting in the American League baseball? What's the ratio of skill to luck in football or basketball? Those are all meaningful questions. You can ask to what extent the teams regressed towards the mean, or players, for that matter, regressed toward the mean in their performance.

Regression toward the mean is very real. It's all over the place. It's an extremely useful prediction tool. The short answer to this question is that there seems to be about a 70/30 skill/luck ratio. Performance in the IARPA tournament is about 70% skill, 30% luck, judging by the amount of regression we observed toward the mean among forecasters in general.

Interestingly we don't observe very much regression at all amongst superforecasters, which defies the laws of statistics. I think I know why -- because there is a countervailing causal force. People really like to be recognized as superforecasters when they move from one year to the next. They want to preserve that status. They have an opportunity to work with other superforecasters and teams. There is a whole series of things that are happening that are elevating the performance of superforecasters and insulating them from regression toward the mean to a substantial degree. Forecasters in general do regress toward the mean. The very talented ones tend to move downward and the worst ones tend to move upward, just purely on the basis of the laws of chance.

Julia: You can, to some extent, tease out those separate channels?

Phillip: You sure can. There are very standard statistical tools for doing that.

Julia: Then does that actually address the question about how much of their success comes down to the amount of time they spent investigating the question?

Phillip: Well, I have a pretty hard quantitative answer to your first question. I don't have a hard quantitative answer to your second question, because it was very difficult to estimate the amount of time and cognitive effort the forecasters were expending per question. I think that's a very important topic for future work. I think it's certainly true that the superforecasters did work hard.

I think it's also true, and this is relevant to the first commandment in the 10 commandments for superforecasters, that they are really good at triage.

Julia: How so?

Phillip: If we ask superforecasters a question for which there are some very readily available, high-quality, public substitutes for their own predictions... They would do what many people would think to be cheating and they would simply say, "I'm going to go for that."

Instead of trying to figure out what the probability of Hilary Clinton being the next President of the United States is, they might turn to a site like Nate Silver's FiveThirtyEight and look at the various analyses that Nate Silver has done, the poll analyses and the trend analyses. Rather than doing them all themselves, simply count on other people. You count on the futures market, you count on some knowledgeable website to do it for you.

The superforecasters often look at two or three different things and average them. If they don't have a lot of time and there are readily available, high-quality, public data sources, superforecasters might well decide not to invest very much time in a question and instead invest their effort in questions where there are no high-quality public data sources and where analytical effort is more likely to pay off.

Julia: At the workshops that my organization runs we tend to have prediction markets. Which is partly just a fun game and partly a way to train the habit of making predictions and putting numbers on your uncertainty. We have this saying: that cheating is technique. Which is exactly that.

To some extent I think the need to say this speaks to the difference between following a process and trying to get the right answer. Those two things often line up. Sometimes I think people think that the virtuous thing to do is to try to figure it out and use their own head or something like that. But actually the boundaries of our brains, now-a-days, extend far beyond what's immediately available to us. That information can be incredibly valuable.

Phillip: I'm not saying the superforecasters will always defer to these sources. I think they are well aware that they'd better have some darn good reasons for [for example] disagreeing with futures markets. Incidentally, prediction markets were an important part of our work as well. They were one of the bench marks against which superforecasters were compared.

Julia: You talked about the one most important thing that superforecasters do. Do you want to go into anymore of the nitty gritty of the methods that your team used to win?

Phillip: Do you want me to?

Julia: Well, to the extent that they are relevant to the advice that you would give to people wanting to become superforecasters. Maybe it's not relevant?

Phillip: Well, the first big takeaway, I think I mentioned, the practical significance is -- it's a necessary but not a sufficient condition. You have to believe it's possible to develop this as a skill. If you don't think it's possible because probabilities don't even apply to these things, or if you don't think it's worth your time and effort to do it because you don't really need to know the difference between a 60/40 bet and a 40/60 bet in your line of work... If you fall into those categories then you probably won't want to invest effort in becoming a superforecaster. If you think that improved granularity and accuracy of probability judgments will improve your life and your work performance, and if you think that it's doable within the task domain that you work in, then the payoffs are potentially very large.

Julia: It's ironic, actually, that the thing you need to have in place to have the motivation to improve your forecasting skill is this belief. Which is not a very intuitive belief because it doesn't ... Just looking at it intuitively, 60/40 vs 55/45 doesn't feel like that big a difference. Our intuition doesn't easily appreciate how much that adds up to over time.

So in order to override your intuitive judgment, that 60/40 vs 45/55 shouldn't really matter, you need to already have the skill that's important for being a good superforecaster.

Phillip: Right, yup. In Kahneman's great book *Thinking Fast and Slow*, there's an explanation for that phenomenon of why it's not so intuitive. That is that people really have a hard time distinguishing shades of maybe. That the amount of the weight, that a probability of 30% gets relative to a probability of 70%, the difference is not that large. When you move from 90% to 100% or from 10% to 0%, it has a huge effect on decision making.

We call it the probability weighting function in prospect theory. It's really flat in the middle, which means that people have a very hard time seeing the value of becoming increasingly granular in that middle zone. It's a perceptual blind spot and it's an expensive one. Great poker players and great investors are able to capitalize on that.

Julia: That brings me to a question I had about this idea of incremental updating, of taking in new information and letting it adjust your probability estimates even if it only adjusts your estimation a little bit. Which as you explain in the book is an important thing that superforecasters are doing that most people don't do.

This is something that I advocate. I talk a lot about Bayes Rule. I do videos on Bayes Rule and Bayesian thinking and this is the main thing that I try to get people to do. I refer to it sometimes as snowflakes of evidence -- that these little snowflakes might not shift you that much on their own, but over time they can accumulate enough collective weight to break a tree branch.

I do believe that's important, but I worry sometimes that because our intuition is not that fine grained about degrees of confidence, that if people try to follow the advice about incrementally updating, that might just cause them to over update. Basically that they will get these little pieces of evidence and that will just sway them far too much.

The superforecasters seem to have avoided that problem, and I'm curious how they did it.

Phillip: This is a topic I would probably go on for hours if you let me.

There are of course two types of scenarios people could make in this situation. They could fall prey to belief perseverance, and they could fail to update when they should have. Or they can fall prey to what people in finance call excessive volatility, in that they over update in response to every little bit of flaky news.

People manage to make both mistakes. Not simultaneously, but to alternate between them quite rapidly. We're more likely to make the belief perseverance mistake when we have strong ideological and emotional priors on an issue and we just interpret all the contradictory evidence into our schema. We're more likely to over update in financial markets where we don't have any great ego investment in a stock, but we're really jittery about getting second-guessed by other investors, so we over adjust there.

There are situations where both can occur. We call this "riding the error-balancing bicycle". For every cognitive bias you can make in life there is an equal and offsetting bias. You could be over adjusting, or you can under-adjust your beliefs in response to evidence. You can be over confident, or under confident and so forth. It's kind of a dialectic you have to manage.

Superforecasters for various reasons are very good at it. I think the best way to learn to be good at it ... Well, there are two ways to do it. One is by doing training exercises and engaging in conversations like this in which we think about it. But I think there are limits to how far you can go with that, too, with the didactic approach. I think you need to get on the bike and try to ride it.

We talk about the philosopher Michael Polanyi and how he distinguished between tacit and explicit knowledge. The explicit knowledge would be trying to learn how to ride a bicycle by reading a textbook on Newtonian mechanics. That would give you the guidelines for how to the cosines and the coefficients, so you don't fall off when you go into a curve on a bicycle. Whereas the tacit knowledge approach is to get on the bicycle and start trying to ride it.

The analogy of that here is getting into a forecasting tournament and start making judgments. The Newtonian manual would be analogous to our probability, reasoning, training, and the book has a lot of guidelines as well. There is no substitute for going to forecasting tournaments and actually engaging in this process.

I'll just plug [www.goodjudgment.com](http://www.goodjudgment.com) because that's the new forecasting website that is the sequel to the Good Judgment Project. People can visit that. There are lots of questions that are posted and there will be some interesting competitions emerging. It will be some work that will involve work with the government and there will be other work that involves more stuff on domestic politics and sports and things like that.

Julia: Great. We'll link to that on the podcast website as well.

Phillip: Thank you.

Julia: The need for tacit knowledge and therefore the need for practice is the next thing I wanted to talk about. I think your track record of showing improvement in forecasting ability is quite compelling. I thought it was pretty cool to see even Daniel Kahneman, who's been a leader of the heuristics and biases field, but also a pretty prominent skeptic of our ability to improve our built-in biases... To see even him saying "Phil Tetlock's track record is very compelling evidence, and maybe suggests that I have been overly skeptical." That was very cool.

I like the way that you described yourself as an optimistic skeptic. That's how I think of myself as well, about this domain.

That brings me to my question: how much do you think practice in a domain like The Good Judgment Project's domain of political and economic forecasting ... How much your improvement in that domain transfers to other domains? Would we expect the superforecasters in the Good Judgment Project to also be better calibrated and to use the same kinds of skills in predictions about how people in their life will react? Or about whether a purchase will make them happy, etc.?

Phillip: I don't really have a lot of evidence in the Good Judgment data set to directly answer your question, but I do have some knowledge of the research literature on transferred training. It gives us grounds for being pessimistic. If you want to learn how to become a well calibrated weather forecaster, or a high discrimination granular well calibrated forecaster in meteorology, make meteorological predictions. Don't count on playing poker to get you through.

It really is very important, I think, to focus on ... I do believe in subject matter expertise. I think I say that in the book, that I'm sometimes identified with this kind of "know-nothing" position, of "expertise is bunk", but that's bunk. I don't think expertise is bunk.

Julia: Then maybe we could talk a little more about why you have been an optimistic skeptic in contrast to, say, Kahneman's pessimistic skepticism. My sense -- and correct me if I'm wrong -- is that even before the Good Judgment Project you thought it was at least plausible, feasible, to improve people's built-in biases or you wouldn't have tried at all.

I have two theories about why Kahneman is more pessimistic than we are, or at least was more pessimistic, and probably still is to a large extent. They are, first -- that the research in the heuristics and biases literature so far has looked at relatively short-term simple interventions to see if people can overcome their biases. Like telling people about a bias. And that hasn't really worked.

I'm not that surprised by that. I wouldn't have expected, a priori, an intervention like that to work.

The hypothesis you were testing is more like “Does intense deliberate practice over time improve people's biases?” which is I think a much more promising hypothesis, and a different one than has been tested in the literature Kahneman was working in.

Phillip: I think deep deliberative practice that's informed by some key concepts like the “error riding bicycle,” and some key heuristics, is the way to go. And the best evidence for effective long-term debiasing of problems would be consistent with that. That's a fairly substantial undertaking.

But our training didn't take all that long in the IARPA tournament. It took only an hour and the effects ramified through the entire forecasting year, which was truly astonishing. I think when people have training, and they start to apply it quickly, it tends to be stickier.

Julia: Do you think that Kahneman ... I suppose I should just ask Kahneman, but do you think that his pessimism was about that kind of deliberative practice?

Phillip: No, well, Danny Kahneman was a colleague of mine at Berkeley and he and Ann have been good friends of my wife, Barbara and I for a long time. We talk about kind of that friendly argument that Barb and Danny had in Chapter ... What is it? Chapter I guess 11 of the book. We talked about scope sensitivity.

This is a hard thing to describe in an audio modality, but in Danny's book he had this... illusion, which is a really, really compelling optical illusion. You see these lines and one line looks longer than the other and you have to pull out a ruler and you have to measure it and you see that those lines are identical in length. Then you take away the ruler and they still look different.

Some perceptual illusions are very [sticky]...

Julia: Yeah. I mean I suppose you have to just ...

Phillip: If you believe biases are like that, if you believe that cognitive biases are analogous to perceptual biases like optical illusions of that sort, which are... virtually impossible for people to overcome other than by long enough rulers or other measurements, objective measurement devices, because our perceptual system can't do it.

... If you believe that cognitive biases are like that, you're going to be very much a pessimist. If you believe that they are more the product of irrational deliberative thought, you're going to be more optimistic that introducing correctives will help.

Julia: Right, or it depends on what you're trying to do when you're trying to overcome biases. If you're trying to get over even the initial impulse to bias, then that would be very difficult if they're similar to visual biases or illusions.

But if you're trying to just make better decisions, then a different way to do that is just to be able to recognize, "Oh, yeah, this is one of those cases where the lines look

different lengths, but they're actually not" -- and to be able to just override your initial system one impulse. Which seems like a fair workaround to me, although it's effortful.

Phillip:

Right, but the scope insensitivity was interesting because it has really quite powerful implications on how accurate you can be as a forecaster. For a lot of people, you ask them the question how likely is Assad to fall from power in the next 6 months, or the next 12 months or the next 18 months? If you ask one individual all those versions of the question, they'll give you somewhat different answers at different time frames.

Imagine you take a group of regular forecasters and you ask 1/3 of them the 6 months question, 1/3 of them the 12 months question, 1/3 of them the 18 months question. Are they going to give you very different answers? The answer is no, by and large. They're going to give you roughly identical answers for 6, 12 and 18 months, which suggest that their answers are not sensitive to temporal scope.

When you do that with superforecasters, you divide superforecasters into 3 groups and you ask one of them the 6 months question, and the other the 12 and the other the 18 months question -- they're not perfectly scope sensitive, but they're much more so.

Now why is that? It's an interesting puzzle. The superforecaster seemed to be thinking in a truly more probabilistic way. Whereas the regular forecaster seemed to be thinking in more of a causal hunch way.

The regular forecasters are saying something like, "Well, let me see. What's the balance of power in Syria here right now? It looks like it may be tilting towards Assad that the Russians are intervening. I'm going to put it at 80/20 he's going to survive. Yeah, it would be 80/20. They're going to 80/20 for 6 months, 12 months or 18 months, right across the board -- because what they're doing is they're translating that kind of balance of power of cognition they have in the brain into a probability, just implied directly.

Whereas the superforecasters, they look at the balance of the power stuff, too, but they're also ... They're more likely to ask themselves a question about, well, how would my answer be different if I were asked about a month or 12 months or 24 months. They spontaneously mutate the question and ask "How would my answer be different if a few parameters in the questions are different?" That's a great way to do sensitivity testing of your positions and it's something that the best superforecasters, I'm not saying all superforecasters do this, but I think the very best ones have a real flair for this. It's truly impressive.

When I talk to very senior investment people in Wall Street, or for that matter, Daniel Kahneman, what's impressive about superforecasters is not that they're hard working. It's that, my god, how could they possibly be scope sensitive.

Julia:

Yeah, it is a very sticky bias. That trait that you're pointing to of the superforecasters has seemed increasingly important to me even for questions that don't... sort of personal choices or emotional questions that don't have an easily objectively measurable right answer.

The tendency that I've seen in some people to ask themselves, "Well, okay, I reacted this way, I reacted with this emotion in this situation, but what if the person had phrased it in different way? Or what if I had been more successful at my job this year or something -- would I have been have felt as resentful of the way they made that comment?" To be able to do that kind of sensitivity test of your own emotional reactions, and discover, "Huh, actually this person's choice wasn't the main determinant of my bad mood -- because I can see that the bad mood could have come from all these other things," for example. It seems like a very valuable skill, that I believe to be very valuable, even though it's harder to demonstrate with accuracy scores.

Phillip: Yeah, it's running little fun experiments in your head, on your own intuitions -- some people do it so much that I think they can do it quite efficiently. It's a mental habit to cultivate.

Julia: Yeah. We're almost out of time, but I wanted to ask you one last question... I studied economics when I was younger. I don't practice it now but I do retain some of the heuristics of economics that I think are particularly valuable. One of them is this question, which..

Well, I suppose one way to express it would be with this old joke about two economists walking down the street, one the younger student and one elderly professor. And the student says, "Oh, hey look, there's a \$20 bill lying in the gutter." The elderly professor immediately says, "No, there isn't. Because if there were, someone would have already picked it up."

Which is kind of a silly joke, but also pointing out a valuable question to ask oneself -- which is if it seems like there's easy money, you should ask yourself why hasn't someone picked it up already?

I guess I'm wondering, with superforecasting and with trainable good judgment in general, it seems like there would be a significant profit incentive to develop that. And I'm wondering why companies in competitive markets haven't already made superforecasting a standard practice.

Phillip: Okay. I love that joke by the way. I think it's a pretty good joke.

A couple of reasons, one of them is more psychological, the other is more sociological. The more sociological reason is that forecasting tournaments tend to disrupt status hierarchies, especially stale status hierarchies.

Okay, I'm 61 years old and I can say I'm an intelligence analyst. I'm the senior intelligence analyst on China. I'm the national intelligence council. I'm the go-to person for the presidential daily briefing on China. I invite influential persons, intelligent astronauts and really important guy.

Someone comes along and they tell you, "Hey, let's have a forecasting tournament here." A forecasting tournament is a level playing field competition. A 25 year old



analyst can offer probability judgments to compete against the 61 year old analyst probability judgments. How I'm going to react to that? Probably roughly the way Tom Friedman is likely to react to the Bill Flack story in the book, very unenthusiastically. That's a big reason, a huge sociological organizational reason why there would be inertia here.

The other thing is a lot of people, as we discussed earlier, are very skeptical that there is room for improvement. Or that even probability just is not possible in these domains. I think this is the first really large scale empirical demonstration that it's possible to improve, that there's probability estimation of real world events that many people previously thought were unquantifiable, and not estimatable on a probability measure. I think it's a big discovery frankly. A \$20 bill that was not discovered because it's covered up quite a bit.

Julia: Right. Yeah, that's a good way to put it. Nice adaptation of the joke.

Well, unfortunately we are over time for this section of the podcast so I will regretfully wrap up this conversation and we'll move on to the Rationally Speaking pick.

Phillip: Okay.

[musical interlude]

Julia: Welcome back. Every episode, we invite our guest to introduce the Rationally Speaking pick of the episode. That's historically been a book, movie, website or something, whatever else tickles their rational fancy -- but this time I'm introducing a new wrinkle, which is that I'm asking our guest to name a book, movie, website, et cetera, that was formative in their thinking, something that shifted them in an interesting way.

With that introduction, Phil, what's your Rationally Speaking pick of this episode?

Phillip: When I was in graduate school in the late 1970s at Yale, I was in psychology but I was very interested in political science. And a book was published in 1976 by Robert Jervis, *Perception in International Politics*. I thought it was just brilliant. I still think it is brilliant. And it persuaded me that political psychology might be a viable field in which to specialize. I think that book had a significant impact on me in a number of ways. It's an old book but it's a good book and I think it's worth reading.

Julia: Excellent. Well, we will link to that on the podcast website as well as to Superforecasting and to the good judgment project website.

Phillip: Okay.

Julia: Phil, thank you so much for joining us. It's been a pleasure having you on the show.

Phillip: Yup, it's really nice talking with you. I hope we have a chance to meet sometime. You're a delightful conversationalist. You have really smart questions. One of the best interviews I had.

Julia: Oh wonderful.

Phillip: A pleasure to talk to you.

Julia: I'm so glad this is audio because I'm sure I'm blushing right now! Thanks again -- and this concludes another episode of the Rationally Speaking podcast. Join us next time for more explorations on the borderlands between reason and nonsense.