

Opinion Science Podcast

Hosted by Andy Luttrell

Why Polls Matter with G. Elliott Morris May 23rd, 2022

Web: <u>http://opinionsciencepodcast.com/</u> Twitter: <u>@OpinionSciPod</u> Facebook: <u>OpinionSciPod</u>

Andrew Kozak:

So it's just a lot of--it's trial and error and you know, this is the most imperfect science out there. We're getting better, but essentially we're trying to predict the future and that creates a lot of, um, room for error.

Andy Luttrell:

It's the one thing you can talk to almost anyone about. "Some weather we're having, huh?" "Did you hear it's supposed to snow again this weekend?" And I think we often take it as a given that we'll know what the weather's going to be like throughout the week. We make plans based on the oracle of a weather forecast...but do we really understand how these things work? I talked to someone who knows.

Andrew Kozak:

My name is Andrew Kozak and I am a TV meteorologist, who has been doing it for just about 18 years. When I was about four years old, we lived in a two-family house. My grandmother lived downstairs, and she had one of those old Zenith TVs that was like a piece of furniture and a VCR hooked up and put in the wizard of Oz. And I don't know, I, once the color kicked in, once she opened the door and she was in munchkin land, I really didn't care about the rest of the movie. It was never a favorite movie of mine. I know a lot of people love it, but to me it was that tornado part in the beginning. Just because it was explained to me that, look, this is a fantasy movie, but there are actual tornadoes, things like that happened. And it just sparked my interest. Oh my God, this is something that's in such a crazy fantasy movie, actually real.

Andy Luttrell:

So Andrew majored in meteorology in college and interned at a TV station in New York City. Since then, he's forecasted the weather all over the place, including in Kansas City, Austin, Memphis, Tulsa...and now he's on Spectrum 1 News where I live in Columbus, OH. Anyhow, I wanted to know—how do you predict the weather?

Andrew Kozak:

On a daily basis, what I usually do is, you know, get up with my cup of coffee. And instead of, you know, everybody reads the news or everybody does a certain, you know, what I do is I start looking at models and data and the latest stuff that comes out while you're sleeping. There's new

data that comes out. I mean, it's just constantly crunching. So I'm looking at some weather models, I'm looking at the overall, even weather discussion that the national weather puts out just to see, you know, these guys have been working overnight just to see what we're seeing.

Andy Luttrell:

And in terms of what these models can account for, as I understand it, there are at least three components. One is just like what's the weather right now?

Andrew Kozak:

I call it NowCasting, what it is now is done by weather balloons and observational data.

Andy Luttrell:

And then we move into more complicated computations.

Andrew Kozak:

...a mixture of the computer model's algorithm, looking at satellite data, looking at a storm in the Pacific Northwest and calculating how long it's going to take to get to Columbus, Ohio where I am or New York.

Andy Luttrell:

And this all gets stacked up against...

Andrew Kozak:

...historical data. In other words, I'm not, today is what the 17th, I am not on, uh, on the 17th going to be able to think that it's going to snow tomorrow in Florida because you know, it knows that.

Andy Luttrell:

And the result of those computer programs, of course, is a perfect description of the weather you will experience for the next 5 - 7 days. Right? Yeah, not so much. I can't tell you the number of times I've heard my dad say, "I thought it was supposed to be nice out today!" Go to YouTube videos about weather forecasting, and you'll see people saying "I wish I could be wrong half the time and still have a job." But my favorite is an episode of Curb Your Enthusiasm where Larry David concocts a grand conspiracy that even when the weather will be great, his local weatherman says it's going to rain just so he can have the golf course to himself.

Larry David:

"It's happened before, weatherman. You know it!"

Andy Luttrell:

The challenge is that the weather is complicated—it's the product of a bunch of factors coming together. Plugging all of this into a predictive model can't get us to certainty, but it gets us to probabilities. And as psychologists know well, people aren't great at thinking probabilistically. We're told there's a "70% chance of rain," and we hear that as "it's going to rain!" But really it means that 70% of the time, when the weather conditions are as they are now, it has ended up

raining. Not only that, but these forecasts are for a given region—a whole plot of geography that gets the same prediction. Put it all together and technically, here's what that forecast means...

Andrew Kozak:

We get what two and a half, three minutes to go on TV and talk. We don't have so much time to explain. So we're very, very picky, especially during severe weather or really important weather to give you the forecast in the most succinct, clear, concise way that we can. So this is, this is a great platform for me to explain. This 70% chance of rain means that there is a 70% chance that your area, that I'm forecasting for the grid, is going to get at least 1/100th of an inch of precipitation. Is it perfect? No. Is it confusing? Probably a little bit.

So let's say I'm just forecasting for Columbus. There is a 20% chance on day five that you will have rain. Now, if by you, I mean you living on the north side and somebody on the south side's watching, well that goes for them too. What it doesn't do is talk about pinpointing where exactly within that grid. It doesn't pinpoint where. When we say 20%, that's where we come in and that's where we explain it.

Andy Luttrell:

By explaining it, he means on TV. You might think the weather app on your phone tells you everything you need to know. But Andrew actually made a pretty compelling case to me that the TV meteorologist still plays an important role. These are folks who can look at these simple, brute force forecasting percentages from official models and convey context to viewers, provide the geographic nuance for what the predictions mean, give a sense of how these weather patterns will move through your area, remind us what a probability even is. But still, it's an uphill battle.

Andrew Kozak:

In our weather community, you know, we have private groups for TV meteorologists, and we always joke, well, what about at my house? Because whenever we do live streaming or live tornado coverage, there'll always be well, what about, you know, what about my house? I remember when I, when I lived in Wichita, Kansas, which is smack in the middle of tornado alley and we were forecasting that day for tornadoes. And I remember somebody had wrote to the station and said, well, I live on the corner of 21st and Rock. Do you, do you think that --? And, and it's like, man, if I could tell you that with, with certainty, that you're going to get a tornado or not going to get a tornado at, at your house, I need to package this, sell it, make millions because nobody can do that. So the perception is if it's not happening to me, it didn't happen. And that's one of the biggest hurdles I think we as meteorologists have, and it puts us in a very unique position because yeah, we work for the news, but we are so different than the reporters and the anchors in what we do. They're out there talking about facts, talking about situations that have happened. They're at meetings for city events or school boards. They're sometimes at crime scenes, they're at all different things that are going on that affect everybody. But they're rooted in video of what's going on, facts about what's going on, interviews with people. With us, it's predicting the future. And nobody can do that with a 100% certainty in any way, shape or form. So perception is a very big hurdle and challenge I think for us. And it's been that way for, since the birth of television weather.

Andy Luttrell:

You're listening to Opinion Science, the show about our opinions, where they come from, and how they change. I'm Andy Luttrell. And you may very well be wondering, "Whoa, what was all that stuff about the weather?" Well, my guest today is G. Elliott Morris. He's a data journalist for the Economist, and he played a key role in building the Economist's forecasts of the 2020 U.S. presidential election. These kinds of forecasts have become a central part of the media landscape in the run-up to elections. Can we predict who will win? And how can we predict it?

It seems to me that it's not all that different from the work of a weather forecaster. If I want to know if it's going to rain in a few days, forecasting models turn to different kinds of data. They take a bunch of measurements of what things are like right now and aggregate them together. Then we can see how those weather patterns have been changing, and stack all of that against historical trends. And how do we forecast elections? Well, Elliott will tell us more in a bit, but basically we can take a bunch of measurements of public opinion at the moment, aggregate them together, look at how they're changing, and stack it up against historical trends. I mean, different models have their own quirks, but that's the gist.

But just like a weather forecast is probabilistic, so are these election forecasts. The good ones, anyway. The forecast isn't that "Biden will win" but that "Biden has some particular probability of winning." And just like a 70% chance of rain means it's still completely possible that it won't rain in your area, as we saw in 2016, election forecasts can go the same way. Election forecasts were calling it for Hillary Clinton, but even if there's a 70% chance that Clinton would win, a Trump victory is not definitively out of the question. And that's what happened. Were the forecasts "wrong"? Well, not any more than a TV meteorologist is wrong on the rare occasion that they predict sunshine and you get hit with a shower.

Anyhow, Elliot Morris has a new book coming out next month called Strength in Numbers: How Polls Work and Why We Need Them. It's about election forecasting, but also a lot more than that. He does a really nice job tracing the history of public opinion polling, taking a critical look at what they can and can't tell us, and arguing why polling is central to our ideas about democracy. It's a great read, and I had a good time getting to know him. So let's get to it!

Andy Luttrell:

I thought a place to start was it's an opportunity to get your personal identification and also talk about what polling means, which is this term pollster, which I have had... It's been like a weird, fuzzy term that gets thrown around, and I wondered if you could clarify like what is a pollster and are you one?

G. Elliott Morris:

So, I'm not a pollster. To answer your question in the reverse order, I am a data journalist, and I guess now an author, so maybe just an empirical writer or however you want to say that; we can talk about what that means too, for the Economist. And I am not a pollster because I do not collect survey data and release it publicly or publish it for a client or something. The crux of the issue is that I'm not actually the one collecting data. Even when I write about polls, I'm writing about other people's polls, other pollsters' data.

The history behind the word pollster is also a little bit interesting. This comes into the book. There's a critic of the polls in 1949. He writes this famous book, his name is Lindsay Rogers, called The Pollsters, and it's meant as a derogatory term, as a spinoff of the word huckster, so he's basically calling pollsters like frauds, and the book is sort of maligning them, and the function of polls in democracy is really... It really is like the main... Yeah. It really is the main force against the polls that we have to reckon with when we argue for their validity, which is what the book tries to do.

Andy Luttrell:

So, I don't think I realized that that was the origin of the term. I have the book on my shelf somewhere. I haven't had a chance to read it yet, but pollsters as a term came through that critique. And is this a case of like the community just embracing and owning that label?

G. Elliott Morris:

Well, I think they had developed their label pollsters before, but it was sort of a very insular usage of the word, and then was popularized by Lindsay Rogers, at least that's the history I know. But it's certainly not a... It was not originally meant as a compliment. Now, you know... Well, the relevant point here is that pollsters are still around, and Lindsay Rogers's complaints have proven through history, I guess, to be either false or irrelevant to people. And so, we use it in a neutral way.

Andy Luttrell:

So, the critical thing is you said pollsters collect the responses, right?

G. Elliott Morris:

Right, so the pollsters collect the data, and will process it using whatever statistical techniques that they need to use to make sure the poll is representative demographically, and then publish the findings either for a private client, like a campaign, like a presidential campaign or a congressman, or they release it publicly. Those are what we know as public pollsters or public polls. And that's what the news media deals with mainly.

Andy Luttrell:

So, if you don't do that, what do you do?

G. Elliott Morris:

I write stories with data. That's really the common tie between all of my work, so I do election forecasting where I'll take the polls and draw trend lines through them and try to predict the future, that very narrow window between wherever we are and the next election and explore the uncertainty in polls. I also just write totally unrelated stories that have nothing to do with U.S. politics in my capacity for The Economist. Polls are a really good source of data. They're not the only data we have, so we can tell all sorts of stories with even other political data, like fundraising data, or endorsements, or anything.

Andy Luttrell:

So, you're kind of like a data liaison. These data are out there, and your job is to sort of use them to tell a story and communicate them in a way that people can make sense of.

G. Elliott Morris:

Yeah, and we do collect some of our own data, and we also do a lot of our own original social science on data that other people have collected, and again, they are mainly... I'm talking about surveys, like the American National Election Survey, or the Congressional Election Survey. I think they recently changed their name. They are no longer... or maybe it's the Cooperative Election Survey. Either way, it's out of Tufts University and Harvard. And you know, we'll do lots of sort of statistical tests of the relationships between variables, what's causing people to think the way they think, what's driving voter behavior in areas among certain people, that sort of thing, what statistical methods can we test and develop for our own forecasting purposes using that data? The questions are sort of limitless.

Yeah, and we do collect some of our own data, but there's really so much good research out there. Liaison is a good word because there's lots of really interesting social science that people do that we don't really need to iterate on, that we just need to report on, and so we do a lot of that also.

Andy Luttrell:

And so, how did you get here? There were moments in the book where there were like turns of phrases where I was like, "This feels like I'm reading political psychology," or like some of the academese that you get in academic writing, but that was not the path you took. And so, I'm just kind of curious like what is it that got you to this place? Why do you do this?

G. Elliott Morris:

So, I started being really interested in presidential elections and voter behavior when I was in college, when I was a second year sophomore student at the University of Texas in 2015, so that's not too long ago, or I don't want to believe it was that long ago. And my professor, a political scientist, public opinion scholar named Chris Wlezien talked to me about this thing called the Iowa Electronic Market, which is like a prediction market. One of the only legal or one of the longest legal prediction markets for election outcomes. They used it for academic purposes, so they say, and I was just super interested in election predictions, so that sort of led me down the path to learning about 538, sort of Nate Silver's methods for predicting elections, which sort of... our conversation with them has evolved quite a bit over the last seven years.

And I just sort of instantly got really interested in the academics of political science. That sort of class I took with Chris was the inflection point. And I ended up in data journalism because The Economist hired me basically before I had a chance to go spend six years getting a political science PhD, which I guess I'm glad I didn't spend six years doing that, but it's still... We still think it might be on the table sometimes.

Andy Luttrell:

Would that have been what you did, is that what you're saying? You would have done that if not for-

G. Elliott Morris:

I would have done that. Yeah. I would have gone to a political science PhD after I graduated. But I was hired when I was still in college to do data journalism, so that's how I ended up where I am.

And again, the good part of working at The Economist is you work with lots of smart people. There's lots of PhDs and social scientists on staff already, so I kind of feel like I get to do a lot of the political science social science, and that comes through in the book, as you say. But you know, obviously I'm missing lots of formal training and stuff, but that's sort of scratching the same itch.

Andy Luttrell:

Were you always a data junky? Is that... Because part of what you do is getting in the weeds with the numbers and doing that kind of programming work. Was that always in the picture?

G. Elliott Morris:

I was interesting. So, to share a little bit more about myself, I'm a triplet. I have two brothers. They're both computer scientists. Now, I wasn't really interested in computer programming, or statistical programming, which is really what I do, statistical programming, until basically the second two years of college, but apparently there's something about my family where that was really interested in math, and statistics, and computers. So, maybe there's sort of some latent desire or push to do this thing.

But no, I was never... I did not go into college thinking I would become a data journalist or a political science. I mean, as is the case with so many students, they learn what they want to do as they learn, and I still think I'm in a learning phase with this thing too, and the great thing about writing is that you get to learn things publicly. You get lots of collaboration from people. Also working with an outlet like The Economist, people are always happy to talk to you about things. Hopefully, I'll be in that learning phase until I die, but that's why I became a journalist. We'll see. I guess the ability to predict these things obviously comes with quite a large margin for error and that-

Andy Luttrell:

You said you didn't get into statistical programming or analysis in late in college, but what did that look like? Was it part of coursework or was this kind of like... I don't know, there's this download R onto my computer and fiddle around?

G. Elliott Morris:

Yeah. It was a download R onto my computer and recreate the 538 election forecast and try to understand these things better. I mean, at that time I thought data science might be in the cards, some sort of computer programming job, not necessarily computer science or architecture, and so I just thought it would be relevant to learn this thing, but I just go so interested in it as I went along, and I took classes in political psychology, so that's sort of where you say the political psychology comes out in the book. My advisor, Bethany Albertson, really pushed me to do political psychology. I'm sure she would want me to be in a PhD program right now, so sorry to her. And you know, I just got really interested in it and when you're trying to understand what makes people tick as someone who doesn't have access to 1,000 face-to-face interviews all the time with a huge research budget, like I didn't when I was an undergrad, quantitative analysis is the natural way to go.

It's also really good for experimentation when we're trying to answer causal questions. Well, this is way out of the... this is super wonky, but yeah, so that's sort of why I gravitated toward the quantitative. And it all just sort of made sense. I ended up reading a book by my now colleague, I

didn't know he worked at The Economist when I read the book, called Big Data, which is all about the data that's being gathered on us by companies, corporations, governments, that will allow us to understand each other, which will drive market and business revolutions. Of course, this book was like... This book was published like six years ago or something, so it's evidently proven to be I think correct in many ways.

And that's just another thing that just sort of clicked, so I think the pieces for what I ended up doing sort of fell in place maybe rather late for me. It was crunch time there for two or three years. But you know, I wanted to be an architectural engineer before I wanted to be a lawyer, before I wanted to be a political science PhD, so yeah, as is the thing, people just sort of figure out what they want to do.

Andy Luttrell:

And the 2016 election would have been right in the midst of this, right? Like-

G. Elliott Morris:

Yeah, so the first forecasting project was a model for the 2016 election. Yeah. That's right.

Andy Luttrell:

And so, you might have thought that this would have been the thing to dash these dreams that were just forming, where you go like, "Oh, this is very cool." And so much of the conversation surrounding polling at the time was not super favorable in the wake of what happened. So, you could have gone through that and went, "Oh, I guess this is all bunk and why am I wasting my time on this? Let's find something else." But that's obviously not what happened. Do you remember at the time, what did that mean to you that these very high-profile election forecasts apparently, based on how they'd been communicated, didn't come to pass?

G. Elliott Morris:

That's funny. I was writing for the student newspaper at the University of Texas when 2016 happened. I think that timeline works out correctly. And so, I had to take on this role of sort of like the poll defender on staff, as like the only political science, political sort of statistician, budding political statistician in the opinion section, and it just... Maybe it's because of the people I was following and talking to online, or like the way that these tools had been explained to me by my professors, but that's just... It just didn't seem quite right to say that they were broken, or wrong, or that they were useless and that we should never trust them again, and that the way we talk about people in political campaigns or otherwise should just fundamentally change because people underestimated the odds that Trump would win. It always made much more sense to me that these tools of measurement of the population have a historical record of misfiring in key ways that teach us about the tool and how can we learn lessons for those things so we can keep using this tool that has a lot of value to democracy?

I mean, it lets you interview America. It lets you talk to people, which is not something that's necessarily easy to do. It's certainly not easy to do in a representative way. And I think those basically are the two themes of the book, so my thinking about the polls was really shaped between 2016 and 2018. When I decided to write the book was in 2018, but yeah, the 2016 election could have gone... I guess it could have gone either way.

Andy Luttrell:

Yeah, so let's get into the book, which I did... I enjoyed quite a lot. And it was sort of a book that I had always in the back of my head wished existed, which is like a very accessible but doesn't shy away from the details discussion of what polls are, how they work, why they work, when they work, and whether we need them or not. So, kudos. I think it was... I really enjoyed reading it.

G. Elliott Morris:

That's a great summary of the book. Do you want to write marketing copy for the book? That was great.

Andy Luttrell:

Pull this right out of the podcast. Go for it.

G. Elliott Morris:

Do you have a transcript?

Andy Luttrell:

So, one of the themes of it that I really connected to, because it's a question that always comes up for me when I think about what polls are and why they exist, is why do we need them. Do they really serve some purpose? And mostly I mean kind of in a sense of these public polls. Like I get it if you're sort of campaigning and I'm trying to identify pockets where I don't have support quite enough yet, and I'm looking for it, but the grand premise of polling, what is it? To you, why are they important?

G. Elliott Morris:

So, there are I guess two... Well, so you just said you get it with the campaigning, so are you asking me journalistically why we need polling for the horse race? Or-

Andy Luttrell:

I just sort of mean-

G. Elliott Morris:

Are you asking the bigger question too? Because depending on what you're asking, I'll give you a different answer.

Andy Luttrell:

Yeah. I get it as a strategic tool, but what I'm more curious about is the more general existence of polls as a thing that we consume, a thing that we report on, and also kind of what I'm getting at is like what it means in the context of democracy, like the philosophy of why polls are important.

G. Elliott Morris:

Yeah, so there's like... Well, so I'll give you a three-pronged answer. I was gonna do four, but four seems like a lot. So, journalistically, I think the easiest way to start is just sort of like the horse race. We need polls for the horse race because they are semi-objective benchmarks of where a horse race stands that anchors coverage to something beyond how editors feel about a race, or

some straw poll that they conducted which is basically how political coverage worked up until the advent of polls in the '40s and '50s. Those things provide empirical bases for our coverage. It means that readers aren't being hoodwinked, basically, and it just gives strong empirical grounding to work that people can trust, basically.

Obviously, that's not the best case for the polls, so the work I've done journalistically is more on issue polls recently, because we're not in an electoral context right now, or I guess we will be in a few months, and that is like what do people care about politically and otherwise? What are the things that are ailing people in their daily lives that we could report on, that readers, government leaders, corporate leaders could read and act upon to help people? And just like what are people thinking at the fundamental level?

If you're a newspaper, you sort of have to talk about people and we want to understand what people are thinking. Not just in that horse race electoral context I was talking about, but also just in like are they doing good? Do they need an economic rescue plan? More recently, do they support intervention in Ukraine against the Russian Army? Do they think that an ambitious primacy-based campaign for democracy worldwide is still something that the United States wants to be doing? These are big questions of government that people should have a say in, so when we report what they say it sort of helps people get what they want out of their government.

And that leads us into the third part, which is that polls are fundamentally a tool for democracy. They were not necessarily developed that way when George Gallup developed the first scientific polls, or I guess popularized the first scientific polls in 1936 through '48, basically, is sort of his arc of development. He started as like a market researcher. He saw the value in polls because he thought, and he writes this down at one point, if we can do it for toothpicks, why can't we do it for people for politics? And so, this was originally a way to sell people something. It was like advertisement market research. And only grew to have this more substantive democratic meaning later on.

I guess the sort of proof is in the pudding, like the fact that the democratic meaning for polls has endured means that there's something there, but also when we interview congresspeople, or when scholars have interviewed congresspeople and asked them like, "What do you want from your constituencies?" Lots of them say, "I want to know how they feel. I want to know what they need. I want to know how they're feeling about upcoming votes." So, congresspeople are inherently interested in what their voters want, so polls can help fill that gap.

They also fill that gap for national representatives, for statewide representatives like presidents and governors who would have a much harder time talking to their constituents than like a congressperson would with... Even now, with a constituency of 600, 700,000 people, that gets pretty tough, but it's impossible for a president. And so, polls sort of help. They just help the people communicate and get what they want or need from their government, or at least they let them communicate what they want and need from their government. Whether or not they actually get the outcome is sort of another question. Lots of other pitfalls.

Andy Luttrell:

You mentioned presidents would have a harder time just sort of gauging intuitively their constituents, but I love the image of Abe Lincoln just sort of like inviting people to the White House and chatting from who knows where, to come in from all over the place, and that was like a pre-poll version of what you're saying, right, is like at the time you go, "Well, how could I possibly know what people want? Well, I'd have to talk to them." And the innovation now is like we could do that much more efficiently and hopefully reduce the bias of just like, "Hey, who wants to come on up to the White House and have a conversation?"

G. Elliott Morris:

Yeah. That's a really good... Yeah, Abraham Lincoln's what do you call it, his public opinion bats. It's also very interesting that he uses the phrase public opinion bat that early. That shows Lincoln was pretty steeped in either the philosophy of I guess democracy and republican governments, and or emerging literature on public opinion. We didn't really have James Bryce's study of American government that gets cited as one of George Gallup's motivations for public polling, his book American Commonwealth, which talks about public opinion and the role that opinion should have in a democracy. It really is the sort of treatise for the pollsters in terms of the democratic meaning.

So, that just shows Lincoln was pretty in tune with these things. I guess as is typical with Lincoln.

Andy Luttrell:

So, that sort of gets me to changes over time, which I think to me, the most stark fact in the book is the difference in how often people picked up the phone in the past versus today, which highlights sort of the evolving challenge of public opinion polling, so I don't remember the numbers off hand, but can you give a sense of what that change over time is and what it has meant for gauging public opinion accurately?

G. Elliott Morris:

So, it's worth starting maybe even a few decades earlier, so when polling was invented by Gallup, there were basically... There was only way really to interview people, which was you went to them. You either go to them physically in the street, or you could sample their addresses and go to their houses, which is sort of the technology that they developed after the street interview method doesn't go very well.

And then Warren Mitofsky invents the sampling procedures for random digit dialing telephone numbers a bit later in the '70s, and when that starts you have response rates close to 70 or 80%. The vast majority of people you called wanted to talk to the pollster. Now, there's not great research on really why that is. The histories of polling suggest that people wanted to take part in this process, this hyper-democratized process of talking to their government. They felt like if George Gallup interviewed them and they were able to give their opinions, then those opinions would matter, which is a particularly hopeful, I think optimistic view for those people, perhaps wrongly so maybe in hindsight. But it meant that people picked up the phone and answered the polls.

The other way to put this is today, instead of having 70 to 80% of people responding to the poll, depending on how you count this you have like 1 or 2%. So, for every person you want to call, you need to call like at least 100. And that means polls are expensive now, especially if you want to do a live caller phone poll. You have to have someone sit there and dial the numbers for cell phones. You can't auto dial the cell phones yet, so that makes polling really expensive. You have to have someone type in-

Andy Luttrell:

Legally, you mean, right?

G. Elliott Morris:

Sorry?

Andy Luttrell:

Legally, you can't do it you're saying? Or just technically can't.

G. Elliott Morris:

Oh, right. Legally, you cannot auto dial cell phones in America. Obviously, like I get 20 spam calls a day, so someone out there is doing this, but the pollsters won't do this. And they want these rules to be changed so that it would make polling cheaper. It's kind of nonsensical. But on the other hand, I don't want any more spam. But anyways, so that makes polls more expensive. It also means that the people who are talking to you are really engaged in politics, which introduces some biases in polls, so you have to correct for those. You have to make sure the people that you're talking to that aren't watching Fox News or CNN or whatever, or reading Washington Post every hour of every day, you have to make sure you weight up those people in your analyses, make sure they're represented carefully.

It also means just like by random selection the types of people you talk to can be really weird, because you've shrunk your sample size down basically by a factor of 100 already in just who you're talking to. So, if you're not gonna have a very high chance of talking to... to use an actual, real-life example, a young Black man from the Midwest, you're gonna have to give that person a lot of weight in the poll once you actually do talk to them to make sure their numbers in the poll is representative of the number of people in the population as a whole. And you could just so happen by the randomness of people, get a person like that who is a Trump supporter, so in the 2016 L.A. Times poll this happens, and they get a poll that's like three or four percentage points too biased toward Hillary Clinton and then switches overnight to be super biased towards Donald Trump because this individual has a very, very high weight in their poll.

And that's just a fact of the statistics that people have developed for public opinion polling to date because of these very, very small response rates.

Andy Luttrell:

Because it's hard enough to sample certain subpopulations, and then it's made even harder by the fact that nobody picks up their phone, so when people come across like one person who suits some demographic category, they then have outsized influence just because they happen to be the one person in that subpopulation that picked up the phone that day. And that's the challenge, right?

That now the gold standard of truly just randomly going through the country and just talking to people on the phone, which still... People will go like, "Well, these surveys are not truly probability samples, right? They're not actually just randomly getting people." But I think what you're saying is we can't do that anymore. Just logistically, we can't do that anymore.

G. Elliott Morris:

Yeah. Well, that brings up a lot of other points from the book too, that all the adjustments that you make to the poll mean that this margin of error that the pollsters are communicating... So, actually, let me take a step back. So, when pollsters communicate their poll, they also will tell you, and this doesn't really get covered nearly enough, that because of this random deviation in who you talk to, like with a 100% percent response rate you're going to get an X percentage point margin of error, that if you just so happened to talk to other people that were slightly more or less representative, you could have instead of 20, maybe like 23% of people who love Coca Cola or something, and I don't know, I don't love Pepsi, so let's just say the number is 5%. Whatever.

So, that's your standard 3 percentage point margin of error for a poll of 1,000 people. It's just one simple formula entirely based on the percentage of the people who support something and the number of people you interview. Now, when your response rate isn't 100% because people aren't like marbles in a bag, they can decide whether or not they want to respond to you, and because the population that pollsters are trying to reach is so increasingly polarized and demographically complex, that means when we are talking or when pollsters are talking to the people and then adjusting their data, they have to make sure it's correct not only on age and gender, but age, and gender, and race, and educational attainment, and like I was saying earlier, political engagement, and some of the interactions between these things. That means there are a lot of different quantities you have to estimate and if you have one or two weird people in that formula, it's gonna throw the entire thing off. And so, pollsters have also developed this thing called the design effect, which means depending on all these interactions and weights you're doing that your margin of error should be even bigger, sometimes like 30 or 40% larger.

And that's before we even talk about the difference between the population that you can sample on the phone or on the web and the sample that's actually going to turn out for elections. So, when we're doing election polling in particular, the researchers have found out that the actual margin of error is like twice as big as the ones that pollsters report. And so, this is sort of the story at the beginning of the book that says how can we better understand this tool and understand why this 2percentage-point margin of error is here so that we can better improve the communication about polls and hopefully decrease the flack, basically, that pollsters get when they have an error in a presidential race of like 2 percentage points or something.

And I mean there's no clear answer. We'll get to that later. That's the end of the book. Helpfully, I make you read the whole thing before you get an answer. But you know, it's a really difficult problem to solve. Basically, you're trying to talk to people and they're not picking up the phone. There's really no great solution to that.

Andy Luttrell:

Yeah. So, this notion in the book has a couple stars next to it in my PDF, which is what you're saying, right? That the margin of error is usually just sampling error, just like, "Oh, we're gonna

be wrong a little bit because we only talk to some subsample of everybody." But sort of the whole premise of the book is like yeah, but that's like just the boring part of the margin of error. There are all these other systematic things that could come into play that push those numbers around. And it reminds me, it just sort of... As you were describing it, there's this great thing that mathematician Steve Strogatz, I was listening to him in an interview, and he was talking about why social science is so difficult, and he's like, "The problem is that if you're a physicist and you want to measure the moon, you can do it because the moon doesn't mind that you're measuring it." In the world of people, when we want to measure people, it's a moving target, right? It's this slippery fish that we can never quite get a handle on because the fact that we want to measure a physical property of something, we could just do it. It's not this moving target.

G. Elliott Morris:

So, the interesting thing here is we've talked about three types of error for the pollsters so far. There's this random sampling error and just the types of people you talk to are different than the types of people you need to talk to along all the demographic attributes, whatever. There's something called coverage error, which is like maybe the types of people who have phones aren't even representative of the population, so you have to find some other way to talk to those people, so sampling and coverage error. There's some error in the processing of your data also, so that is the third type of error. And then there's a fourth type of error which is just measurement error, which is like if you're asking people questions, the way you ask them the questions is going to change their responses. Like if you say, "Do you approve of the president who just had this horrible scandal?" They're going to be like... That's obviously a bad way to ask the question because people are gonna be like, "No, I don't like anyone who likes a scandal." Or, "Do you approve of receiving \$2,000 in the mail from the president? Do you approve of the president?" People are gonna be like, "Yeah, that sounds great. I love the guy."

And then equally, if you're asking people about sensitive questions, then you're gonna get different answers too. So, we know that there are other things, like this is I think what you're describing is sort of similar to social desirability bias, which is if you're asking people their height, people, or at least men, tend to overestimate their height because it's socially desirable to be taller. There's other examples of other appendages on the male body that also come up with surveys a lot. People overestimate how much money they earned, et cetera. And so, there's also... The relevant theory in election polling is it's not socially desirable to be a republican Trump supporter, at least the theory goes, so therefore people have lied about being Trump supporters, and the idea behind this sort of like shy Trump effect is that that explains why the polls have underestimated Trump in 2016 and 2020.

Well, actually, the rules or the explanations are a bit more complicated. It's not about social desirability but this just suggests there's a whole world basically of things that can go wrong in a poll. My question in the book is what can we achieve for democracy if people understand that those things exist.

Andy Luttrell:

Yeah, some of the solutions are in the how we design these things and the mechanics of the polls. I'm curious as a communicator, as a journalist, what could we do, and you give examples of this, but just to sort of spark the conversation, what could we do to convey more openly and transparently and accurately what we can and cannot get from polling numbers?

G. Elliott Morris:

Yeah, so at the end of the book I have a few suggestions for basically increasing public literacy about polling which will hopefully have downstream effects of increasing trust when pollsters say whatever, that 20% of people like Coca Cola or whatever I said earlier. I think was 20%. We know only 5% like Pepsi. That's for sure.

Andy Luttrell:

That is established fact.

G. Elliott Morris:

There's no margin of error on that one. And there's your pull quote. But right, so how can we improve public conversation about the polls? Well, I think there are three really good starting points. The first is that when pollsters publish their polls, their estimates of what share of the country like Coca Cola, they should not only be releasing the margin of sampling error. We know that that three-percentage-point error is too small. We know this empirically because the polls are often wrong more than 3% and also theoretically, because there's both sampling, or because there's four types of error, not just one. There's not just sampling, there's also sampling, and coverage, and non-response, and some measurement error. And the pollsters could really actually be clearer about those things existing. It's not their fault that these things exist, but they could at least... And you know, they'll talk about them in conferences, but in their press reports or press releases in the reports, these things of polls if pollsters just sort of did those basic things.

Second, I think election forecasters and the press covering election forecasts could be more careful in how we talk, and I use we because I also forecast elections, in how we talk about the polls and our estimates. We give probabilistic forecasts for elections. In 2020, The Economist said Joe Biden has a 97% chance of winning the election. It would take a large polling error across states for him to lose. And technically, we were right about that. We said 97% of the time in past elections polling errors have been so small that the person in the lead, Joe Biden, still became president. And that's sort of what happened, but that 97% really makes people optimistic. They think when we say 97% that, "Oh, The Economist is predicting a landslide for Joe Biden," and that's how people reported it, which is not really true. We were saying he was gonna win 340 electoral votes with like 8 percentage points of the vote. That's not technically a landslide depending on how you count it. Whatever.

But if we instead told people a polling error slightly larger than the polling error in 2016 would be enough for Trump to win again, people would say, "Oh, that's not actually so far flung. That threepercentage-point chance or 3% chance is relevant to me. I remember what the polls were like in 2016. Something else could go wrong. Maybe that 3% is not so far off." We could also present our estimates, this is sort of technical now, with like terms of conditional probability. We could say, "Okay, conditioning on there being no bias, Joe Biden has a 97-percentage-point chance. But if we just presume that all the polls are as biased in the same direction towards Biden as they were in 2016, actually his chance of winning is only 54%." And so, people sort of will think of these things differently. You're sort of giving them a scenario to anchor their understanding of the world and their understanding of the accuracy or potential inaccuracy of the polls, and then they won't overreact when the polls are eventually wrong.

I mean, the two-percentage-point margin of error or three-percentage-point margin of error means you should really expect the number in reality to be different than the precise number that the pollsters are reporting just the way statistical distributions work. There's a lot more, again technical, mass in the distribution in the tails and in the small errors to the left and to the right of the point estimate than there are right around that whatever, 20-percentage-point Coca Cola point estimate. And forecasters could be clearer about that.

I think the last thing is that the association for pollsters, the American Association of Public Opinion Research, could also be clearer about what polls are good and what polls are bad, or what methods are being researched and those pollsters that are clearly in violation of basic standards of collecting or reporting data. There are several sources of biases in public opinion research. I think the biggest one right now is ideological biases. There are several polling outfits that will go on Twitter and tweet about how bad the president is, or some of them even echoed conspiracy theories about the outcome of the 2020 election. Not to get too political here. But there's reasons for us to think that those pollsters are also processing their data in ways that might be favorable towards one party or the other, and empirically they are often more biased towards Republicans. So, we have to wonder should AAPOR be saying, "You should be more suspect of these pollsters. Here are our recommendations for how these practices should work," or, "Can you at least share your methodology?" Because both of these pollsters also aren't transparent in how they're even conducting their polls.

And I think those three things, so this public sanctioning, to go in reverse order, of bad pollsters, this emphasis by forecasters and people who report on forecasting of uncertainty and of conditional probability, and this disclosing by the pollsters of the potential error in their data would go a long way toward improving the public conversation about polls and would I hope just prompt people to give polls a second chance when they're so-called wrong, right?

Andy Luttrell:

In some ways to that effect, 2016 was an asset, right? It was an example that we can point to that says like, "Hey, this is how probability works. When I say this, remember it has happened before that these kinds of numbers didn't actually pan out in this way." Like most of the time they're gonna conform to this law of probability... I mean, it's always according to the law of probability, but like to the conclusion that I'm drawing, like most of the time my conclusion is gonna be accurate. But hey, you know, sometimes based on stuff that either we're not accounting for, or just the realities of what a 30% chance means, it's gonna happen.

Maybe another way to couch it is sometimes... Well, people hate the weather, so maybe this isn't a great idea, but you can say, "Sometimes the Weather Channel says it's gonna rain and then it doesn't," right? It even said, "There's a 70% chance it's gonna rain between 1:00 and 2:00." And sometimes it doesn't, and we should expect that, because that's what 70% means. So, part of what I'm hearing you saying is like the language of uncertainty is unfamiliar to lots of people, but if we

could put it in a context of concrete events that demonstrate what uncertainty means, that could go a long way to sort of illustrate how we should use these polling results.

G. Elliott Morris:

Yeah. I think the weather example is kind of interesting. You know, when you get the weather forecast, it's actually telling you that in your tiny area you have a 70% chance of rain. In the immediate vicinity of you, there should be 70% rain or something, and then also the other uncertainty. So, it's not just that the forecast could be wrong. It's also that they can't actually forecast for your exact location. They have to draw a certain boundary around you. That is like the same thing that's happening here with the polls. There are multiple factors driving the uncertainty in the model.

A good way to think about this is that pollsters have inputs. They have, just like weather forecasters do, they can observe the demographic traits of people and what they think, just like a weather forecaster can observe the physical traits of the air today and how much that has changed in whatever target horizon they're gonna give you, and then both of those things go into a model. In the weather forecasting sense, it's like they know 70% of the time in this area when the weather traits have been this, there has been rain. Forecasters know, or pollsters know, that 95% of the time when you observe attitudinal estimates with this composition of people, with these precise weights, collected from this precise way, asked in this way, XYZ, whatever, then you end up with Coca Cola 20%, plus or minus 3, right?

And yeah, so imagine if... I mean, that's a lot of caveats for a political reporter to put in their report on a story but imagine if every news outlet had a page on their website that says, "Here's how polls work," or, "Here's our interview with these people on how polls work," and they linked to it every single time they wrote a story about the poll. There was a small disclaimer and they said, "Here's more about the poll," or whatever. Polls can be more accurate than the single estimate says, blah, blah, blah, read more. That would already go a long way towards increasing literacy among people who pay attention to the polls, and the way to reach people who don't watch the news, I guess, is to do more podcasts where people listen to podcasts, so thanks. Or I guess to do this on cable news, too, where you just sort of have to have segments where you explain how polls work. People need basically to be taught this thing. It's very complex but it's because it's so important both to our narratives of politics, and campaigns, and to outcomes from our democracy that understanding the complexities is very, very important.

And I think reporters should spend more resources on that.

Andy Luttrell:

I wanted to sort of pivot a little bit to talk about forecasting a little more specifically, and you know, I had kind of been for a long time pretty lukewarm on election forecasting. Not that I think it's wrong or bad, I just didn't think I needed it. I was like, "Let's just all vote and then see what happens. Why does everyone need to have this idea ahead of time what's gonna happen?" But in reading the book, the insight that I came to is that actually election forecasting is super important because almost never do we have the opportunity to check a polling estimate against some objective benchmark. The results of an election are like the actual population estimate that we're trying to approximate, whereas in issue polling, we don't always have that. We go like, "Well, our

polls say X% of the population agrees with this statement," but we don't actually have the national number to check that against.

So, it strikes me that predicting elections has actually been maybe the most generative source of innovation in public opinion polling. Does that seem right to you? Is that where the sort of extra methods and all these additional innovations have been coming from?

G. Elliott Morris:

So, in the sense of polls as election forecasts, yes. I mean, George Gallup, I have a book here of lectures from George Gallup in 1949. He talks about polls as forecasts, so it's fine I guess for us to talk about those things, and that's what prompts not the normative criticisms from people like Lindsay Rogers, but the reactions from the press that say, "Oh, we can't trust these people. They're soothsayers. They're hucksters. Whatever. They need to fix their methods and then come back to us." And that's what prompts in 1948, for example, this report by the Social Science Research Council that says, "Switch to probability sampling of households instead of interviewing on the streets," and they switch to probability sampling. That's what makes George Gallup use probability sampling.

That's what causes after 2016 pollsters to start weighting their data, or at least many pollsters, not all of them, to start weighting their data by the interaction between race and education, not just race. So, we know that non-college-educated white working class people were sort of underrepresented in surveys, and now they are presumably not, at least. Presumably. So, in that sense, electoral misfires have caused real methodological changes by the polls. The flip side of this is that we need... We really do need, and by we I mean like democracy and consumers of the polling data, need reminders that these are just estimates. After 2008 and 2012, the election forecasters, here not the pollsters, the people predicting the elections using the public polls, were pretty liberal in claiming their victory, their ability to predict election outcomes. Nate Silver routinely said his models got 50 states right. He was right about the election.

And that framing of right versus wrong is sort of damaging to the pollsters who don't predict things as yes or no binary outcomes. And by the way, Nate Silver, and I, and election forecasters also aren't doing things in binary ways. We've talked about these probabilistic predictions already and how we're exploring uncertainty. So, the right versus wrong framing here can also be pretty damaging and it reminds us as the communicators of the polling data to relearn all of these sources for error in the polling data so that we can keep communicating about the polls in a methodologically sound, responsible way.

Andy Luttrell:

Yeah. You mentioned also a little bit ago about... When we talked about election forecasting, you're like, "Well, if we just say that polls are the tool for doing that, then we can have this conversation." But it's not simply like I'm just looking at a few snapshots and making a guess. I at least know in hearing you talk about the model that you helped develop for The Economist to forecast the 2020 election, there's more to it than just we're reading in and summarizing polling information, right? It's not just that, so what else is going on? How fancy can we get in terms of

accounting for more information to make these predictions? What are the kinds of additional inputs into a forecasting model beyond just public opinion survey results?

G. Elliott Morris:

So, at The Economist, we have a forecasting model that has like three sources of information, basically. We have the polls and those get added on to what we call a prior prediction, and we use the language of Bayesian statistics here where you have a starting point, the prior, that you put the data on top of that inform your model. And that prior is made up of our prediction using fundamentals data, that's political and economic fundamentals, so what we do is we collect data on economic growth from today, whatever day we're forecasting, all the way back to 1948, and we look at the historical correlation between whatever economic growth we're observing and the outcome of the presidential election.

So, the idea here is when there's a recession, presidents get punished, and when there's economic growth, they stay in the office longer. We also look at how predictive presidents' approval ratings tend to be, so we assume some error here in saying that the polls are actually good measurements of a president's approval rating, but again, the theory here is when presidents are more popular, they win more votes in November. When they're unpopular, they get kicked out. So, we train regression models on these historical observations and then we add the polls on top of that.

Now, that fundamentals prediction serves basically to anchor the election forecast. It says, "You're in an environment where the incumbent president hypothetically said he hates Pepsi, so he's more popular because of all the people who like Coca Cola," or whatever. And then you add the trial heat polls on top of that, and that can prevent your model from straying into territory that says the incumbent president is gonna lose by 10 or 15 percentage points, to go back to reality, in like a hyper-polarized political time. And it also allows us to sort of... Again, similar to this MRP thing I was telling you about, to expressly put our knowledge of the world, of our social science-rooted understanding of how people are making their decisions and what shapes election outcomes, it allows us to expressly put that into the model and then update whatever we think that is with the actual data, with the actual observations of the horse race, so the model can figure out, "Oh, what we're anchoring to is wrong. Actually, people are making decisions off of X, Y, or Z factor. Maybe it's COVID. Maybe it's whatever to use the 2020 election and you need to ignore the prior and instead rely on the data and add uncertainty," right?

All that stuff happens inside the model. And that's contrary to if you just had the polls. You said, "Okay. Well, we can only really trust the polls here," and what you're doing there is you're saying there's no other information we could gather that would be instructive of the election outcome in some way. That's not particularly true, so we use the other model that lets us formalize our theories of voter behavior and of uncertainty in the polls and all the other factors.

Andy Luttrell:

So, it's like even before you've asked a single person how they would vote, there's already some sense of how this thing is gonna go based on the current state of the world and what that meant for election outcomes in the past. I'm just curious how, like if that's all we had, how good would we be? How accurate would we be? You I guess know from those regressions, like how much variance is accounted for just by this stuff?

G. Elliott Morris:

Well, we also have a neat little variable in our model at The Economist, which no one else uses, which captures polarization. So, we think... That's sort of the other underlying factor here is we think in polarized political times that elections are more certain. Empirically, they are, because there are fewer swing voters, so we have a long history or a recent history of elections in the United States where presidents win between 53 and 47 percentage points of the two party popular vote, whereas like in 1964, Lyndon Johnson was winning 60% of the two party vote for president, right? That doesn't happen anymore because there just aren't that many swing voters. There are more people who hate the president, basically, and will never vote for them, than like 40%. So, they'll never reach that, so we can tell the model like actually, before we even collect the economic data, because we know that we're in pretty polarized times, only forecast roughly between 40 and 60 percentage points.

And the track record of the model after we include that variable tells us that our historical error is about as good as the polls. It's like one-and-a-half to two percentage points on the incumbent president's vote share. That's the standard error. That means that your margin of error is closer to like four percentage points. That's basically what you get with the polls on election day, but it's much more accurate early on in the election cycle because it's forward looking, because we sort of used some machine learning techniques to optimize the regression, not to bore you with all the ML stuff, and because people change their minds.

When you ask, the literal question that gets asked in an election poll is if the election were today, who would you vote for. Not if the election were in November, right? People can't forecast that. And what's happening today can change between today and November, basically, so that's why there's more error in the polls, and that's why we rely on the fundamentals prior before election day comes around. And why it has tended to be pretty accurate.

Now, the thing I was immediately thinking of when you asked how accurate this thing was is in 2020, our fundamentals prior on election day had Biden up four points of the two-party vote. Our polls had him up like eight-and-a-half to nine percentage points. So, evidently there's something about our theories of elections that can break with the actual observations, so it helps to hedge between those two things. I mean, really what we're doing is social science research. As per usual with social science research, you don't want to trust observations just willy nilly that are totally divorced from your theories of in this case elections, or in their case whatever they're studying. You want to like understand why there's a difference between those things, and then either you say, "Oh, those differences are captured in the error term of our model in the case of election forecasting, that's why there's uncertainty," or you try to adjust the model. In our case, it's really hard to adjust the live forecast on the fly, and tell people why we're changing our assumptions, so we tend to do the former thing, but there's a lot of room... I'll say there's a lot of room in election forecasting for more statistically sound ways to adjust the forecast over time. It's just it's been really hard both computationally and theoretically, but there's lots of really smart people working on this now, so we've made lots of steps in how the model is working, basically. The statistics of how models of election outcomes work even though they're roughly as accurate as they used to be.

Andy Luttrell:

Nice. Well, this has been great. Just by way of wrapping up, if you want to give a quick pitch for the book, what it is, what people can expect, and when it comes out, so this episode will come out before the book comes out, so let people know when they can get it.

G. Elliott Morris:

So, the book is available for preorder now on W.W. Norton's website or my website, gelliottmorris.com. It comes out July 12th. It is about basically what our entire conversation has been about. The people who listened to this conversation will probably read the book and be like, "Oh, I already knew that," but there's lots of story in the book that we haven't talked about here, obviously, because it's like 70,000 words. So, anyway, yeah, the book is about how polls work, why there's all these different sources for error, the story about how pollsters have tried to combat all those sources of error, and why we need better public debate, public conversation about polls, so that we can sort of use them to their fullest extent, which is elevating the voice of the people.

Lots of people, just by way of one final anecdote, lots of people come up to me and they're like, "Oh, people can lie with the polls." They throw around the lie, it's easy to lie with statistics thing all the time, which is a quote from this Swedish mathematician, Andre Dunkels. The actual full quote here is it's easy to lie with statistics, but it's harder to tell the truth without them. And so, I typically tell the same thing to people when they're like, "Oh, there's uncertainty in the polling data." It's like yes, actually, there is lots of uncertainty in the polling data. Maybe even more than two or three percentage points that we observe in elections. But without them, we would know almost nothing about what the people want, how they behave, and so that's roughly why we need them.

And after I write about that quote in the book, there's like eight more chapters, so that explains why, and that's what people can look forward to.

Andy Luttrell:

Great. Well, it was a great read, and it was great to get to talk to you, so thanks for being on the show.

G. Elliott Morris:

Yeah. Thanks very much, Andy.

Andy Luttrell:

Alright, that'll do it for another episode of Opinion Science. Big thanks to G. Elliott Morris for taking the time to chat about his work and upcoming book. And even though he made it seem like we covered the whole book in this episode of the podcast, that's definitely not the case. There's a bunch of great stuff in the book that we didn't get to. If you're into public opinion, politics, or journalism, you should definitely pick it up. You'll find a preorder link to the book and a link to Elliott's website in the show notes.

Also thanks to Andrew Kozak for filling me in on the finer points of weather forecasting and the importance of local meteorologists. Be sure to follow him on social media for weather-related tidbits.

Okeedoke—make sure you're subscribed to Opinion Science so you don't miss an episode. There's some fun news about this summer's episodes of the podcast coming next week. Stay tuned! That's it for now. See you in a bit for more Opinion Science! Buh bye...