

TO S2 E3 - SCRIPT

**CLIP [Barack Obama, Election Night 2008]:** *Hello Chicago!*

**RAFFI VO:** The 2008 U.S. Presidential election was a historic one. But not just in the way you might be thinking.

**CLIP [Obama]:** *It's been a long time coming but tonight, because of what we did on this day, in this election, at this defining moment, change has come to America.*

**RAFFI VO:** Barack Obama was a once-in-a-generation candidate. He gave captivating speeches. He promised to reach across the aisle, and mend a divided country.

**CLIP [Obama]:** *That we have never been just a collection of individuals or a collection of red states and blue states. We are, and always will be, the United States of America.*

**RAFFI VO:** All of this was truly historic. But there was this other revolution — in the way his campaign was run.

**CLIP [Obama]:** *To the best campaign team ever assembled in the history of politics, you made this happen, and I am forever grateful.*

**RAFFI VO:** And the digital strategies of the 2008 Obama campaign *really would* make history, forever changing political campaigns. Not just in the internet era, but in the era of social media.

**CLIP [NEWS REENACTMENT]:** *So, after Barack Obama announced his candidacy, it was clear that he was the online frontrunner. Like, going on the blogs, on YouTube, on Facebook, on MySpace, you could just tell that he was gonna get the nomination.*

**RAFFI VO:** Obama wasn't the first politician to use data, of course. Polls, surveys, census data, and even things like Nielsen ratings and demographic trends have always been leveraged by campaigns looking to reach voters — and win elections. But Obama's two campaigns took data and analytics to a whole new level.

**CLIP [CNN YT, Obama 2012]:** *The Obama campaign is widely believed to have had the most sophisticated data mining operation.*

**RAFFI VO:** Political campaigns now are massive data operations.

**CLIP [CNN YT 0:00-0:08]:** *Meet Donald Trump's mindreaders, quietly crunching away five thousand pieces of data about every American adult.*

**RAFFI VO: In the US, candidate teams interface with state and national parties, as well as with firms who provide data services to campaigns — as clients.**

**CLIP [CNN YT 0:37-0:44]:** *Their big data goes beyond your voter profile, or even your magazine subscription. This is micro-targeting.*

**CLIP [CNN YT]:** *Analysts say campaigns do this by going to commercial data warehouses, which compile information on your buying habits, like what magazines you subscribe to, maybe information about where and how you like to travel...*

**RAFFI VO: And this new ecosystem brings along some...unique challenges.**

**CLIP [CNN YT, DNC Hack]:** *This just in on CNN: Russian hackers managed to infiltrate the computer network at the Democratic National Committee and what information did they apparently zero in on?*

**RAFFI VO: The 2016 US Presidential election featured a ton of data mischief and malfeasance. The DNC was hacked by Russia...**

**CLIP [CNN YT, DNC Hack]:** *Perpetrators are affiliated with the Russian government, and the goal was essentially to gather intelligence.*

**RAFFI VO: And other info was stolen as well, as top party officials fell victim to phishing scams.**

**CLIP [CNN YT 5:31-5:36; 5:52-6:01]:**

*Wolf Blitzer: All of your personal emails were hacked by the Russians, and then released through WikiLeaks.*

*John Podesta: Well, they're looking at the incidences of hacking, uh, what occurred, how much damage it did...*

**RAFFI VO: In the wake of this, and after the election loss, I was hired, in 2017, to be the first-ever CTO of the Democratic National Committee. On the one hand, they wanted a bona fide technologist to solve this big cybersecurity problem.**

**But it's clear that they also wanted to change the culture *around* technology at party headquarters. Campaigns were becoming more technically sophisticated, but the thing is: campaigns are pop-ups. After the election's over, all that digital infrastructure just... disappears. Then-chair Tom Perez hired me because he wanted a way to integrate some of this advanced technology into the permanent workflow of the national party, so it could benefit not just Presidential candidates but all candidates, up and down the ballot, into the future.**

At the time I was hired, the so-called “tech team” at the DNC was like four people. But I expanded my team to over fifty — the largest in the building. We overhauled systems that, at the time, were almost a decade old. And we established relationships with tech people and tech companies that endure today.

I also learned a ton about, what I guess you would call Washington politics? But I won't really get into that too much here.

But in the two plus years I spent at the DNC, I got to meet a ton of people working at the intersection of data science and politics. And I want you to hear from *them*. Some worked alongside me at the DNC. Some are former campaign veterans now working at their own data firms. Others are political strategists who've made a pivot away from politics.

But they're all gonna help us talk through what political campaigns are doing with your data.

From Emerson Collective, I'm Raffi Krikorian, and this is Technically Optimistic.

[THEME MUSIC]

**RAFFI VO:** In 2012, Dan Wagner was the chief analytics officer on the Obama campaign.

**CLIP [DAN WAGNER]:** *I'm a data science practitioner and in my previous life I worked on this guy's campaign...*

**RAFFI VO:** He led a team of fifty or so analysts working out of the Chicago headquarters, in a windowless, computer-filled room that came to be known as “the cave.”

**CLIP [DAN WAGNER]:** *In 2012, my, my job was to manage a crew of 54 people that was primarily responsible for leveraging this, this kind of huge, incredible store of new data to create the, the first individual-level personalized campaign.*

**RAFFI VO:** After the campaign, he founded the data science company Civis Analytics, where he's the CEO. In my time at the DNC, I got to know Dan and he liked me, I think, because we both understood the power of data-driven tech in politics.

So I called him up to talk about what his firm, and campaigns in general, even do with all this data.

Where does it all start?

**DAN WAGNER:** The first piece of information about voters is called the voter file, which is really the single source of truth that campaigns and the party committees have collected about voters.

**RAFFI VO:** The voter file is ground zero for campaigns, and it contains some basic information about the voting public. Dan likes to give this little piece of history.

**DAN WAGNER:** An interesting tale is that the first kind of main analytics political organizer was Abraham Lincoln, who in early county elections in Illinois ran a very simple campaign where he told his volunteers to go out and collect opinion about everybody in the county. And they came back and they said, this is the name of the person. They're undecided. They support me. They support the other guy. And they took all those cards and they shoved them into a box.

And then a week before election day, Abraham Lincoln opened the box and he said, these are the people who are undecided. You go and talk to them about issues. These are the people who are with us and might not vote. Go talk to them to vote. And these are the people who don't care and aren't going to vote anyway, so just ignore them. That was the voter file. It was a box of people with information about their turnout history and their opinion that he organized into a single place to organize all their communication. Today, the voter file is that, but on a computer.

**TARA MCGOWAN:** So I'm going to, I'm going to put my, my former hat on as a political operative and strategist I am no longer.

**RAFFI VO:** Tara McGowan is a former political strategist, and a digital media expert. In 2019, she founded Courier Newsroom, and is currently focused on supporting local, digital journalism. But in her former life as a strategist, Tara knew all the ins and outs of the voter file.

**TARA MCGOWAN:** Which is like the list of voters, the primary kind of tool we're talking about here, you know typically has an individual's name, their address, their zip code, hopefully an email — although email is tricky — hopefully a cell phone, but that's even less prevalent, but it's getting a little bit better, and whether they're registered or not, and then if they voted and in what elections, never who they voted for, but their party ID, whether they've declared it or you are modeling for that.

**RAFFI VO:** And we'll talk a little more about what Tara means by “modeling” in a bit. Here's Dan Wagner again.

**DAN WAGNER:** So the database itself is gathered from state parties who get it from secretaries of state or boards of election. It's maintained as public record. It is the people who are registered to vote and their turnout history for the time that they have been registered to vote, and their address. And in some voting rights states, there's information about race as well, that the government uses to essentially measure equal participation.

**RAFFI: Uh, birthday, gender...**

**DAN WAGNER:** Birthdate, sometimes. In a state like Pennsylvania, you'll get birthdate. In a place like New Hampshire, you might not. Often, gender you don't have, but you can typically infer it from the name. So it's not required information in some states, but you can, you can infer it.

**RAFFI: And something similar, assuming on ethnicity.**

**DAN WAGNER:** Something similar on ethnicity, yeah.

**MAX WOOD:** Yeah, so, well, my name is Max.

**RAFFI VO: Max Wood was also on the data team for Obama in 2012.**

**MAX WOOD:** Basically, if you are a random person running for office, in most places, the county government will say, here's your voter file of everybody registered to vote in the district you're running for on like a CD-ROM or something.

Most campaigns have access to a much cleaner version through some kind of CRM product that their state party makes available to them, whether they're a Republican or Democrat, and they're able to see all this information.

**RAFFI VO: CRM stands for Customer Relationship Management, and it's a software tool that businesses use to manage the data they have on their customers.**

**Political campaigns use CRMs to manage data on voters.**

**For instance, at the DNC, the CRM was called the VAN — the Voter Activation Network. And it was the interface we used all the time to view voter data. I can't emphasize how important this is to Democratic campaigning. Volunteers, organizers, campaign managers — everyone *lived* in the VAN.**

**But the voter file? That's only one source of data that campaigns might get access to.**

**MAX WOOD:** I run a small company called Indigo that helps Democrats and progressive advocacy organizations generate more effective content for the people they wanna reach.

**RAFFI VO: Max is gonna help us understand how the data collected from the voter file can help with what we mentioned before: *modeling*.**

**MAX WOOD:** So we have all these traits about the candidates. And then we have a bunch of traits about the people who cast those votes. This included like their demographics their best estimates we can come up with that their socioeconomic status information sort of derived from all of that. So like the household composition. Are you a Democrat registered Democrat who lives with a registered Republican, you know, are you married to somebody of the opposite party? And so all that interacts in this like complicated set of nested arrays to predict the election result. And then we basically would take those models and run them on an individual.

**RAFFI VO:** Every data point that a campaign has on a person can be used to make a prediction about how that person might vote. How much importance to give each data point, and how to interpret certain relationships between data points — that’s the kind of work done by data analysts like Max and Dan.

Here’s an example of data modeling: They might design a formula to synthesize many data points on a given individual, add in some assumptions about which data points matter more than others, and spit out something called a “persuasion score” — that’s a number between 1 and 100 representing the likelihood that someone might be convinced to vote for your candidate. That’s just one of the things Dan Wagner’s team built on the Obama campaign in 2012.

**But a model alone can only take you so far.**

**MAX WOOD:** So, you know, there can be statistical problems with making conclusions about individuals from aggregate data.

**RAFFI VO:** When I was at the DNC my deputy was an amazing woman named Lindsey Schuh Cortes. Now, she’s the CEO of a company called TargetSmart.

**LINDSEY SCHUH CORTES:** We are a full service data and data strategy company. So TargetSmart does everything from knitting together the 50 Individual state voter files, appending a bunch of consumer data to that, appending a bunch of model data to that, and then shipping that out to the world.

**RAFFI VO:** In providing services to a campaign, Lindsey’s firm deals with much more than just voter file data. Hearing about how that data is organized can help us understand how data firms go beyond merely modeling to *targeting*.

**LINDSEY SCHUH CORTES:** So we have access to a bunch of consumer data sets sourced from all over, right? Like everything from, you know, the Cat Fancy Magazine of the world to, you know, modeled household income. And so once you look at that, it kind of expands that world of what was on the voter registration file, which, usually it's first name, last name, date of birth, like address...and then it suddenly expands it by about 1,500 additional columns.

**RAFFI VO: Holy crap.**

**LINDSEY SCHUH CORTES:** Sometimes more, depending on, you know, which data sets that we mix and match with it. So, 263 million Americans by about 1,700 columns wide is what that file looks like.

**RAFFI VO: Okay 1,700 columns. What kind of stuff is in all those columns?**

**LINDSEY SCHUH CORTES:** So it's, it's a lot of things, right? Like anything that we think or that our customers tell us could be useful in targeting. Likely household income. Marital status, right? Like, how many other people are in the house, right? Like how many other registered voters are in the house. Some of these columns we derive, some of them we acquire, and then consumer data — we need to know who the Volvo driving soccer moms are. And so you would go and find consumer data to see if you could find who the Volvo drivers who have kids who, you know, purchase soccer equipment who have a minivan. All of those things to try to find a soccer mom or a NASCAR dad, right? Because all of that helps to try to make the organizations who are using the data better at reaching out to individuals and having authentic conversations, help them figure out which issues an individual voter cares about or their membership cares about, so they can actually talk to them about things that matter.

**RAFFI VO: That's targeting. So, modeling helps identify certain audiences, targeting is the process of figuring out how to reach those audiences, specifically — and most effectively.**

**DAN WAGNER:** So step one, collect the data.

**RAFFI VO: Dan Wagner again.**

**DAN WAGNER:** Step two, clean up the data. Step three is to measure people's political beliefs typically through survey research or active voter contact. Step four is to use that information to segment people into different groups, and you can segment them based on some of the demographics that we talked about, or by political beliefs that you create from some of these different representative surveys of voters. Once you have those segments created, step five is to personalize your message to those different segments. And ideally try and talk to those different segments within TV, social networks, et cetera, based on what message you think is going to resonate with those different segments.

So you're gonna go step by step gathering information, measuring political beliefs about them, segmenting them. And then personalizing those messages to where you believe those people are in the specific distribution of media channels where they are. Then, at the end, you're going to measure that over time with whether or not you changed people's minds, yes or no, and whether or not you influenced voting behavior to get to 51 percent or not. And then it becomes

this recurring feedback cycle from gathering data, to measuring, to segmenting, to communication, and then back again to understand how effective your marketing campaign is from day zero to the election and back again.

**RAFFI VO: Tara McGowan talks about this same feedback loop in running targeted ad campaigns.**

**TARA MCGOWAN:** I did run really massive digital advertising programs. We primarily ran three different kinds of advertising programs. Persuasion programs. So programs that were meant to put factual information in front of specific audiences of voters to inform and educate them about positions of candidates on specific issues, to move their support or opposition on an issue and/or candidate very specifically. That is the goal of a program like that. And then you need a mechanism to be able to measure that difference from before you start the program, and then after.

And so, we would always run randomized control trials, just like medical trials, where you create a placebo group of a representative sample of an audience. And then the treatment group, you can have multiple treatment groups. So you can try different messages or messaging formats. And then you, you know, pre-survey them before the treatment, they then get the treatment, the placebo group gets a placebo treatment, and then after the treatment is done, you survey again a representative sample of both and you can measure the difference. So that is, I think, the gold standard still to this day of how to measure persuasion impact on a program.

You start with the audience you want to move. [laugh] And so, you figure out, okay, who is this audience? What do they care about most? Where do they live? And, and you build a universe of these people, and a lot of people use models. You figure out what issues they care about, where they stand on them, and you basically match-make. You match-make factual information and content through ads or news or other information formats that tell a true story. And you can test those different messages and ads in this process through these different treatment groups, and you find the one that moves that type of audience the most. And that's where you scale your advertising dollars, targeting that audience at an optimal frequency many, many times, because you can't hear any message once to believe it, especially in this information environment. And then, you know, you, you move people.

**DAN WAGNER:** So we talked about it a little bit earlier: there's the voter file and there's first party data.

**RAFFI VO: Dan Wagner again.**

**DAN WAGNER:** The voter file is the publicly available voter registration information and then turnout history of a person. The first party information is gonna be the result of some type of campaign's communication to you and the information that they collect about you. I send you an email. You register your email. You put your phone number in for text communication.



**LINDSEY SCHUH CORTES:** And so what first party data is in the political space, is all of the data that you get when you send somebody to door knock.

**RAFFI VO:** Lindsey Schuh Cortes.

**LINDSEY SCHUH CORTES:** So like I knock on Raffi's door. So one, he opens the door. That's a data point. Two, he tells me I am likely to vote for this candidate. So that's another data point and three, he says, yes, I'm going to turn out on election day. So all of those data points, all of those questions, all of those things that happened because we reached out to Raffi are first party data.

**DAN WAGNER:** That is then stored against the voter file or it's stored in some other database if you can't match it against the voter file.

**RAFFI VO:** The process of taking first party data, which comes from direct outreach to voters, and consumer data, which can come from corporations or other data brokers, and correlating it back to the info in the voter file — this is known as “matching,” as Dan Wagner explains.

**DAN WAGNER:** What matching means — let's say I have, I have a campaign and that campaign has a list of voters, a list of volunteers, a list of donors, a list of text message people, a list of people in a spreadsheet from five years ago that came to an event.

I have them in different databases, but I don't exactly know who's who between those. And what we invest in, which is critically important, is I want to know exactly who is the same person. Because if I have somebody on my email list, say my email list is 10 million people, I need to know exactly whether or not somebody early voted in the last election or didn't. I need to know exactly whether or not somebody is Spanish speaking or not. And what we do is basically try and join those together to say Rafi Krikorian on my voter file is the same person or not as Raffi K. or R. Krikorian on my email list. Anytime you do matching, it's probabilistic, so I can say this person is the same as this person with a different level of accuracy.

**RAFFI VO:** A lot of consumer data is supposedly anonymized, meaning data that's collected, for example, from your phone's GPS, is stored *without* personally identifying information, like your name or home address. But location data, that shows someone winding up at the same house every night, often gives a clear indication of what this person's home address might be. And if you compare that with the voter file, which *does* list a home address, then you have formed a powerful connection.

Through this matching process, data scientists can wind up with shockingly detailed profiles on people.

**RAFFI VO: So how do you figure out which attributes actually matter?**

**LINDSEY SCHUH CORTES:** So we listen to data scientists across our ecosystem to tell us: Let's try to find this data because we think it'll be more predictive.

**RAFFI VO: Again, Lindsey Schuh Cortes.**

**LINDSEY SCHUH CORTES:** Like marital status, again, like highly predictive in a lot of these models. Presence of Children in the household. We don't have data on people who are younger than 18, of course, but you know, having kids in the household means that somebody might be more likely to care about education issues. So it's trying to figure out what is predictive of the types of things you're looking at.

**Raffi: So, have you ever looked yourself up? If you look me up like, like, can you tell what car I'm driving right now? Like, would that be something in there?**

**LINDSEY SCHUH CORTES:** Probably. I'm not going to do this though. Uh, right? [LAFF] Like, definitely. Yeah, totally.

**RAFFI: I'm not trying to do this on air, but could you tell what magazines I subscribe to?**

**Lindsey Schuh Cortes:** Yep, I could do that. And then I could probably tell you your last five addresses you lived at.

**RAFFI: Can you, can you tell my Instagram followers?**

**LINDSEY SCHUH CORTES:** No.

**RAFFI: Okay so now, so why not? Like, where is, where is this boundary?**

**LINDSEY SCHUH CORTES:** Yep. So Instagram is one of those companies that does not sell who follows whom, right? Like that kind of relational graph. Like we look to figure out if there's ways to, you know, find Twitter handles and match those at the individual level so I could know what your Twitter handle is, right? So I could tweet at you, but trying to find all of your followers, that's just data that, you know, isn't available to purchase.

**RAFFI VO: I asked Max Wood the same question: Given all the data that is being swept up on me, can political campaigns see my social media data?**

**MAX WOOD:** Yeah, that's a great question. Social media is a great question. It depends on the service. And they've all gone through waves of what they make accessible through their API. So

I think, you know, 10 years ago, you could have used the Facebook API to figure out who's following the Barack Obama page. And then you could have done work on your end to try to match those profile identifiers of public profiles to the voter file. And it would have been imperfect, but you could have tried.

That's like in no way possible anymore. And that's generally been the trend is services have seen how upset people get when there's a sense that their privacy is being violated for the benefit of political parties or political campaigns.

**RAFFI VO: Mm-hmm.**

**MAX WOOD:** More so than I think people get upset about corporations getting access to their data.

**RAFFI VO: Which is like generally ironic, right? Like in the grand scheme of things.**

**MAX WOOD:** Yeah, yeah. Yeah. Yeah people are willing to like compromise their privacy I think for like consumer conveniences and like to get better Netflix recommendations or something but something about a campaign using that data to like more effectively like bullshit you really gets people's haunches up and maybe understandably but it's interesting.

**TARA MCGOWAN:** We're all constantly giving away. All of this data for free to massive corporations and platforms that we love to hate.

**RAFFI VO: Again, Tara McGowan.**

**TARA MCGOWAN:** But we're doing it because the value that we're getting in return is clearly higher to us. And I think that I'm speaking up for most people. I think they would agree.

And yet that also means that like, we're signing away rights to our ideas, our content, our photos, that we don't really know the extent of.

**RAFFI VO: The way that these campaign veterans talk about the *tradeoffs* we make all the time — giving up some privacy for convenience, using a personalization algorithm rather than generic outreach — these aren't decisions that are unique to campaigns. In fact, you can start to see these political campaigns as kind of a microcosm of the *entire\_data* economy. All the same features, and all the same problems.**

**We're gonna talk more about how campaigns became this way after a short break.**

[MIDROLL]

**RAFFI VO: Welcome back to Technically Optimistic. I've been talking with a bunch of heavy hitters from the world of campaign data, including Tara McGowan.**

**Tara McGowan:** was a journalist at 60 Minutes and Frontline, and I wanted to go into politics after I covered the 2008 election. I was hired as a digital producer on President Obama's re-election campaign, and I was essentially tasked with helping to build an in-house video production team that would produce videos, you know, stories of voters across the country who were impacted by issues and, and policies. But, the distribution of that video was entirely on the Internet. And so it was constantly figuring out what is the utility of these videos? You know, this is back in 2012 too, right? So this is all kind of horizontal videos for Facebook at the time.

**RAFFI VO: Tara had become a digital strategist on the campaign, one of about fifty or so people with that title. And what a digital strategist does? Well...I'll just let her explain it.**

**TARA MCGOWAN:** It really just means like, people who focus entirely on how do you communicate or how do you drive specific actions among specific audiences on the internet and social media and mobile. Everything's kind of digital now, right? Like more than 50 percent of TV people watch on their television sets is actually digital streaming apps and streaming content, and that will only increase until it hits close to 100%. So, yeah, where it started out as sort of like social media and quote, unquote, new media, we called it, back in the early days, now I think it's, it's just anything and everything that you reach on your phone with a wifi connection.

**RAFFI VO: Unlike the analog world, on social media you get immediate analytics feedback.**

**TARA MCGOWAN:** When you are pushing out content or information on a digital channel, whether it's through Instagram or TikTok, there are various levels of data that you get back from these distribution channels, about everything from how many people were reached by that piece of content, how many people viewed that video that you published, how many people viewed to the end of that video you published, how many people opened your email, how many people clicked on the links inside the email that you sent, right? There's so much data and this is all enormously valuable for optimization purposes, for actually getting better and smarter at achieving your goals, whatever they are. So more people open your emails, optimizing the placement of your buttons and the colors in your email so more people donate or click on a link.

**RAFFI VO: So, on the one hand, Tara is saying that she's gonna let data inform her political outreach, down to the color and placement of a "CLICK HERE" button.**

**TARA MCGOWAN:** If you run A B tests with the same exact email, but different colors of the "donate" button, you're going to know really quickly which color leads to the most donations if there's a clear winner and then you're going to send that email to everybody else.

**RAFFI VO:** But, on the other hand, she believes there's such a thing as relying too much on the data, and not enough on political instincts.

**TARA MCGOWAN:** I think that data is so important and valuable and it is a tool, not a strategy. And I think that's really important. The data is not going to tell me what message is going to deliver these people to vote. That also has to be part of the strategy. Data is imperfect. We know this from polling. Polling is imperfect. And I think when people rely just entirely on what data or polling tells them, they miss a lot.

**RAFFI VO:** Dan Wagner, Obama's chief analytics officer in 2012, has also seen the importance of balancing data with good old-fashioned face-to-face politicking.

**DAN WAGNER:** I mean in 2012, it was a blend between a massively organized field operation that literally involved the work of hundreds of thousands of volunteers knocking on doors, etc. Now, Facebook has a lot more information about somebody's political beliefs because it's reflected in what they've said about themselves on the internet. And so the, the kind of thrust of political communication has moved from a blend of offline and online much more to online communication through the platforms.

**TARA MCGOWAN:** Most of these platforms have privacy policies in place, right? So especially when you're an advertiser, you are essentially boosting content on Meta on Facebook or Instagram to a target audience, and you're getting information back, you're not getting any personal information back.

**RAFFI VO:** In other words, there is an advertising engine built into Meta platforms like Facebook and Instagram.

**These platforms have their own targeting system, so if you want to advertise there, you're gonna be locked in. And that's what Meta wants — for advertisers to rely on them exclusively.**

**Facebook comes up with categories or topics that they think a user would be interested in, based on their activity on the site. Frances Haugen talked with me about this on last week's episode.**

**FRANCES HAUGEN:** And so they invest whole teams of people to say, let's figure out how to distill what we know about someone into a smaller amount of data. And one of the things that would fit within there is the topicality. These are kinda like the subjects of what they think you're interested in. And those topics might be extremely precise. It might be even like the name of a song you like, it might be more abstract, like affinity with African American content. You know, they don't say that you're African American. They just say you have an affinity for this kind of content. They have this whole continuum of abstract concepts to very, very precise ones. And

so that topic model ends up influencing what you get to see or don't get to see, but you don't necessarily know that Facebook has put you in these buckets and that the world that you get to witness is influenced by them.

**RAFFI VO: Meta still allows advertisers to use some version of these categories to do targeting of ads on the platform. But, as Tara points out, things used to be even wilder.**

**TARA MCGOWAN:** The platforms have pretty strict rules in terms of personal identifiable information. You know, Meta certainly used to be much more open about that. And now it's much more closed.

**RAFFI VO: Things have changed since the Obama 2012 reelection campaign. In part, that was due to a very public scandal that broke in 2018, involving the firm Cambridge Analytica.**

**CLIP [Cambridge Analytica]:** *Cambridge Analytica can help you run more cost effective campaigns that reach more voters and deliver more votes...*

**CLIP [Channel 4 (UK) YT]:** *Essentially, the pitch was that we were going to combine microtargeting, but bring on board new constructs from psychology, so that we wouldn't just be targeting you as a voter, we'd be targeting you as a personality.*

**CLIP [The Guardian YT]:** *They boast that they won the election for Trump, accessing a vast pool of data, the names and email addresses for 230 million Americans. For each, they could access thousands of layers of personal information and craft a message to persuade you, individually, that Donald Trump should be your next president.*

**RAFFI VO: Using an app that looked and functioned like a personality quiz, Cambridge Analytica was able to harvest data on tens of millions of Facebook users, and use that data to do targeted, political messaging. This included information about a user's social graph — in other words, who a person's Facebook friends were.**

**CLIP [The Guardian YT]:** *They had apps on Facebook that were given special permission to harvest data not just from the person who used the app or joined the app, but also it would then go into their entire friend network and pull out their data out as well. . . . Things like status updates, likes, in some cases private messages.*

**TARA MCGOWAN:** The thing that always comes up for me when I hear about it, when I'm reminded about Cambridge Analytica is that that took away a tool that was really valuable on the 2012 Obama campaign, which was targeted share on Facebook, for those of you who remember that. A very smart, savvy way of being able to get first party data from folks, on Facebook, that loophole got closed pretty abruptly after.

**RAFFI VO: Wait, wait, you have to explain Targeted Share, just for like one second.**

**TARA MCGOWAN:** I know. I knew you were going to do that. It enabled advertisers, like if they had a voter right? that they wanted, you could actually access all of the data of their friends to then reach them as well by name. So it was just, I mean, it was a privacy nightmare, but Dan Wagner and the smart folks in the cave figured out how to leverage this to do really smart communications programs using Meta. It was totally legal and totally permissible by the platform at the point in time.

**DAN WAGNER:** This is a really interesting story.

**RAFFI VO: Here's Dan Wagner again.**

**DAN WAGNER:** Because there's kind of two conversations about Cambridge Analytica that I think are kind of super interesting. You know, so, so what they did is they created a kind of fake survey, for people to fill out with some kind of, you know, psychographic things that kind of made people mad and they, and they filled it out. And then there was a little thing at the bottom that said, "Hey, are you cool with us taking your information about you and your, your friends and kind of like doing stuff with it?" And of course nobody read it. And they were able to use a loophole and, you know, Facebook's, you know, permissioning kind of stole information about people and their first party contacts.

And in the same way that you do a voter file model, they took that survey and they, using those patterns present from somebody's social profile and their friends, they extrapolated out those kind of critical psychographic features to identify what people were functionally vulnerable to different negative messages. Um, that was bad.

**RAFFI VO: The thing is, as Tara pointed out... what Cambridge Analytica did, and got in trouble for, had some similarities to what the Obama 2012 campaign was doing.**

**As detailed in 2013 reporting by Jim Rutenberg in the New York Times, visitors to the Obama campaign website could login via Facebook, where they would be prompted to grant the campaign permission to scan their friends lists, their photos and other personal information.**

**The campaign could then do matching against the voter file, giving them a Facebook-enhanced view of your real-life social network. The Times article features quotes from people who worked on the campaign, saying that Facebook knew this was going on, but they didn't do anything about it. Which is similar to accusations made against Facebook in the Cambridge Analytica scandal.**

**But according to Dan Wagner, it's not clear how valuable all this social graph information would even really be to a campaign.**

**Dan Wagner:** I think a lot of the the actual effectiveness of that was, was overblown because I'm not sure they were able to use that information at scale effectively.

**RAFFI VO:** It's hard to evaluate whether Dan is right about that, or not. Cambridge Analytica was working on behalf of the Trump campaign in 2016, which of course did win their election. But we don't know the extent to which this targeting worked.

**Fallout from the scandal did put heightened scrutiny on Meta's data policies, and prompted a series of promises from CEO Mark Zuckerberg, and other Meta officials, about how things would change.**

**CLIP: [Zuckerberg testimony]** *When we learned in 2015 that Cambridge Analytica had bought data from an app developer on Facebook that people had shared it with, we did take action. We took down the app and we demanded that both the app developer and Cambridge Analytica delete and stop using any data that they had. They told us that they did this. In retrospect, it was clearly a mistake to believe them.*

**DAN WAGNER:** Nevertheless, like, it was really creepy. I mean, they, they took data using a loophole in the social platform of you and your friends, and then basically used that to identify social vulnerabilities and target people based on it. That's really creepy. I mean the platforms have really locked that down since then.

**RAFFI VO:** But creepiness persists. Because even if Meta might have closed things up on their platforms, because they got caught, out there in the world, data still flows scary free.

**MAX WOOD:** I mean, I mean, I could just talk for hours though about what you *can* get.

**RAFFI VO:** That's Max Wood again...

**MAX WOOD:** I mean, it's truly wild, things that you would think shouldn't be gettable but actually are, right?

**RAFFI VO RT:** Wait, so, like, what else are you getting? Or where else do you get stuff?

**MAX WOOD:** So there's vendors that companies like Experian that kind of gather a lot of this data and clean it up and make it really easy to access. And then historically the big source of this data and that tends to be the source of the most kind of reliable and meaty data are like loan applications, credit card applications, things that where you have to fill out a lot of information.

**RAFFI VO:** Of course.



**MAX WOOD:** Okay, so TV, right? There's laws against cable companies, for example, selling data on what one individual person watched because cable companies have that data through your set-top box. But there's exceptions to the law. There's like clever workarounds. And so a machine learning system would be able to pick up patterns in like what you view, and they were able to do that. People are willing to get very creepy. They just don't want to get in trouble or make people mad. So, I mean, folks, maybe remember Congress passed a like a bill where capable like ISP like Internet service providers could sell your browser history if they had it.

**RAFFI VO:** It's true. A 2017 regulation made it possible for ISPs to collect and share personal data on users largely without obtaining consent. According to a [2021 report](#) by the US Federal Trade Commission, even though some ISPs claim not to sell consumers' personal data, they do allow it to be used, transferred, and monetized by others, and they hide any disclosures about these practices in the fine print of their privacy policies.

**MAX WOOD:** The creepiest piece of data that exists, I think, out on the market right now is location data, live location data.

**RAFFI:** Wait, wait, live location data?

**MAX WOOD:** Yeah, and so you have to be using an app on your phone that has permission to access your location data and is always on and can then send it to a server. Things like, I'm not trying to throw Candy Crush under the bus. I don't know if Candy Crush grabs your location data, but apps of that genre that are ubiquitous, that's why they exist. That's their business model. Bail bonds, companies that end up not getting the bail returned by the court, like hire bounty hunters and they get this like Candy Crush location data and they hunt down people. It's truly wild. And corporations, like marketing departments, will find these companies that aggregate and sell this data and say, "Hey, we want the names of everybody who was at Coachella because we're going to try to market to them about this lifestyle product that appeals to like millennial bohemian hipsters or whatever."

**RAFFI:** That's wild.

**MAX WOOD:** That's very easy to get.

**RAFFI VO:** Campaigns aren't just interested in collecting data on voters in order to do targeted messaging, or make accurate predictions. In the U.S., at least, campaigns are also concerned with fundraising. Like very concerned.

**Don't get me wrong —on Election Day, they definitely want your vote. But in the leadup to Election Day, they also want your money.**

**And if you've ever given money to a political candidate, and watched your email inbox or your text messages explode, then you know what I'm talking about.**

**MAX WOOD:** So, people will know this if you donate to a campaign, your life is ruined, you're constantly getting messages from campaigns because you're in the system in this very public way.

The other thing that happens here is campaigns sell their donor lists to other campaigns. They sell their opt-in text message lists to other campaigns. Campaign C did not get your permission. They bought the list from Campaign B. Campaign B did not get your permission. They bought the list from Campaign A. Campaign A got your permission, but then you promptly unsubscribed and said stop because you thought it was annoying. But you didn't unsubscribe and say stop to campaign B or campaign C. And so that's one major way your data ends up in the system.

**RAFFI VO: And why would campaigns sell data to one another?**

**RAFFI VO: Tara McGowan again.**

**TARA MCGOWAN:** A lot of organizations can't afford to have an in house data team. I mean, that's really important. And so, you can be a tiny, scrappy organization with two to three people running small programs and use Facebook ad managers targeting of geography and interest level and things like that and run a decent program. So there are these very user friendly tools that are available if you're, you know, buying ads for a couple thousand dollars on Facebook or Instagram. But in general, the more sophisticated it is, you also need the internal capacity to know what to do with it.

**RAFFI VO: In other words, the ability of a campaign or any organization to run their own data analytics operation, and not just rely on Facebook, all depends on the resources they have at their disposal.**

**LINDSEY SCHUH CORTEZ:** I think people are using more data and here's why.

**RAFFI: Lindsey Schuh Cortes again.**

**LINDSEY SCHUH CORTEZ:** Resources and budgets are always tight, especially on the campaign sides, right? Like, people are raising money, but it's never enough money. And so, in order to be the most efficient, you need to find exactly the right people. You don't want to waste time. You don't want to waste money. Reaching out to people and trying to persuade them to vote for you, if they're never going to vote for you. You're wasting time. You're wasting resources. And so especially as organizations have become more savvy, and they have over time, campaigns, candidates have figured out that, like, data is the best way to make your resources more efficient.

**RAFFI VO: And Lindsey came up with one way for organizations to get more access, by pooling their resources together. It's called the Democratic Data Exchange, or DDX, for short.**

**LINDSEY SCHUH CORTEZ:** So the Democratic Data Exchange is a for-profit company that takes in first party data.

**RAFFI VO: First party data, as a reminder, is data that comes from direct outreach to individuals.**

**LINDSEY SCHUH CORTEZ:** And so all of that first party data lives at that organization, right? It lives in whatever CRM or database they have it in. Every organization has their own little database full of these or sometimes very big databases, right? Full of this type of data. So what DDX does is it takes all of that data in, anonymizes who sent it to them, and then makes it available for organizations, because they put data in, they're able to take some data back out. And so the concept of DDX is exchanging data, in an anonymous way, to allow for better outreach across the ecosystem.

**RAFFI VO: DDX's mission is to increase the amount of data available to Democratic and progressive organizations, so that they can target more accurately. And DDX runs on the practice of blind exchanging.**

**LINDSEY SCHUH CORTEZ:** So the way that the blind exchange happens is, you know, Organization A pushes in 10 data points, Organization B pushes in 15 data points. We know who those data points are about. We know they're both about Raffi. Because we, at the individual level, have identified Raffi.

But the thing that's anonymized or blinded is who sent that data in and the reason it has to be blinded is because of campaign finance regulations and campaign finance regulations as it relates to data that flows across the ecosystem, there is a hard break between the campaign coordinated side — so that's candidates, state parties, party committees — and then the independent expenditure side. So like PACs, super PACs, advocacy groups. Those two sides of that firewall can't exchange data. By law, nothing of value can cross that firewall. And so, what happens at DDX is they take in data from both sides of that firewall, both the campaign coordinated side and the independent expenditure side, and then scrub out who put that data in so it doesn't show up for other people who are looking for that data.

It's still tracked within DDX, of course. But the organizations that go in and want to pull out data never know who put that data point in.

**RAFFI VO: One believer in DDX's mission was Tara McGowan. Tara sees more data sharing, and more data in general, as a way to do better *targeting*.**

**TARA MCGOWAN:** There's so much noise, right? And there's so much information and free information and content flowing on the internet and on social. In order to break through that noise, you have to meet people with something that is relevant and compelling to them. And politics is the least relevant and compelling thing to the vast majority of Americans in this country. And so you really do have to, as we say, weave the vegetables in. You have to package it by starting with something they care about and then backing into it. And so, but I think it's really, really important because that's actually how you're going to build trust.

**RAFFI VO RT:** **And speaking of trust, I wanted to ask Dan Wagner: If voters knew how much data harvesting was going on in political campaigns, wouldn't they find it weird?**

**DAN WAGNER:** In some ways yes, in some ways no. There's really nothing new about this. And so there's some kind of understanding that this is going on. I think where it has maybe become weird is that the platforms do have so much information and then they can, you know, target you based on that information online in the same way that a traditional consumer marketing campaign can, and that can certainly be weird.

The other part that I think people find weird and creepy is not necessarily the targeting, it's the you kind of wake up and you're kind of being blasted by text messages and calls, etc. And you're kind of like, where has my information gone? And that, in many cases, has been inappropriate and I think people should reasonably say, you know, that that's kind of bullshit.

**RAFFI VO:** **Inappropriate in part because all this consumer data being collected by campaigns and the data firms working for campaigns actually doesn't amount to much. That's not my opinion, that's Dan Wagner's.**

**DAN WAGNER:** .In our experience, consumer data has not been typically helpful as a predictor of political behavior or political belief. So if I have somebody's consumer history and they like cat food or they bought x, y, z car or didn't buy x, y, z car, and you kind of match that to somebody's voter profile . . . the additional incremental statistical value and prediction power of whether or not somebody bought cat food or didn't or this car or that car really doesn't provide a lot of value. And so I think there's a lot of kind of like conversation about oh my god, there's this consumer data and I think people may overpay for it because in practice I'm not sure it provides any that much value beyond the information that a campaign is collecting through its normal activities.

**RAFFI VO:** **This was kind of surprising to hear. It costs a lot of *money* to go out and acquire purchase history, credit information, and other consumer data. If there's not a ton of value in it, why do it?**

**MAX WOOD:** Um, it's like the paper clips problem. You just continue asking for more data budget until, you know, the world no longer exists. Um.

**RAFFI VO: Here's what Max Wood had to say about all the consumer data being obtained by political analysts.**

**MAX WOOD:** None of it matters. It's ridiculous. The data is a mess, first of all. All this data, I said something like maybe 100 million people are leaking their location, but that's not everybody. And so when you have a model that only knows real facts about half of the people that you're trying to train on and has like guesses for the other half or missing data for the other half, that feature is not going to be that valuable to you. So that's one thing.

**RAFFI VO: Another thing, according to Max, is the fact that the voter files themselves are in pretty rough shape.**

**MAX WOOD:** The voter files are really wacky. If you like get married and change your name, they often will keep that as a second record. There's not like...you think public data, all the public agencies are cooperating. That's not the case. If you die, state government doesn't notify the county registrar, you know? They just have to wait until you don't vote for three elections and don't respond to their mail in many cases.

**RAFFI VO: But Tara disagrees with Dan and Max on this. In my conversation with her, she immediately thought up situations where, if she had access to more consumer data, she could use it to do targeted messaging.**

**TARA MCGOWAN:** If you had purchase level data about certain things, then you can, you know, use that to inform the content strategy and the targeting strategy of an audience that, you know, purchased an electric vehicle. They obviously support electric vehicles. If you have to get them communication about that, that would be likely very valuable.

You need to buy a bigger house because you have, you know, you've had two kids in the last five years and you've outgrown your house, but you're scared to leave it because interest rates are through the roof. Hey, by the way, like there are all of these houses for sale in your neighborhood, like a partnership with real estate data, right? And then say. There is a brand new 10,000 tax credit for second time home buyers. And like, that's a way of doing it so you're able to get that message across about the policy.

**RAFFI VO: But on the other hand, Max Wood points out a problem with this approach.**

**MAX WOOD:** I mean, like, this is the thing also, it varies across all the jurisdictions where this data is managed, the standards. And you're like building a model and making decisions that treat all this data as homogenous in quality. And it's not. And so the more like creepy and invasive and like esoteric your like data is, the sort of worse it is.

**Raffi: But then why do people want so much of it? Why is there so much effort then spent like purchasing this data, layering it in and all this stuff?**

**MAX WOOD:** So like targeted use cases make sense. This is just my personal view, I think we want to do things really well, and we want to do things really intelligently. The people who are in control of budgets are not the people munging through JSON records of all this data. And they want to like micro-target using the latest information. And so they approve budget to buy all this data that has limited predictive power and actually maybe makes the models worse because it adds a bunch of noise, but they're also like, this is really fun to play with. And maybe they think there'll be a use case for it later.

**RAFFI: So. Gaining access to consumer data. Does it actually matter? Or does it just make things messier? I guess we could say that even the experts disagree here. But, the thing is, either way, it's *your* data, and your privacy on the line.**

**Whether it's actually helping target voters, or if it's just due to momentum to get as much data as possible, campaigns are out here purchasing your consumer activity. And I don't know how aware people are about this.**

**Okay Lindsey, let's think about an average voter. Or an average consumer for a second. Odds are, when a person buys a car, or a burrito, or a Cat Fancy magazine, they're not thinking about that action turning into a piece of data. That's then fed into this complicated system and visible to all sorts of people working on campaigns. How do you feel about this? Does it make you uncomfortable that there's such a lack of awareness? Or do you think, like, people should be more aware, because that's just how the world works?**

**LINDSEY SCHUH CORTEZ:** Yeah. I mean, so for me, it feels kind of just, this is how the world works, right? Like that is, that is a point, that transaction point. I use my credit card when I bought Cat Fancy. That data is going somewhere, right? Like, do I actually care? Kinda no. Does it mean that I'm gonna get more cat, uh, rescue ads? Maybe. I'm kinda fine with that.

But thinking about, like, my folks, for instance, who, you know, are in their seventies, I think they're a lot less into that idea, where it's like, I don't want you to sell it. I don't want anybody else to know that I'm purchasing Cat Fancy Magazine.

And maybe it's just because, like, there is more of our lives on the internet as you get younger and younger. Like, my son, for instance, like, downloads all the apps on his phone. I'm like, don't do that. Don't do that. They can track. They can get more data. And he's like, why do I care if they know I play Candy Crush?

I'm like, they know how long you play Candy Crush. They know, like, all of those types of things. And then they're going to link it back to our house.

For me, it's: why do you need access to my photos? Why do you need access to my location? Why do you want access to all of these things? Why would an app like this need to track my photos, right? Or have access to my camera?

I think it's interesting that privacy legislation in all of the 50 states is now kind of catching up to this.

**RAFFI VO: And indeed, the fact that we don't have a national data privacy policy is a huge deal. There have been a number of different attempts at this, even just this past April. But so far, they've all come up short. We've already talked a little about this this season, and we're gonna talk a lot more about it in later episodes.**

**But let's think about what Lindsey was saying. Right now in the U.S., there's no protection in place against an iPhone game collecting GPS data and selling it. But lawmakers are spending their time instead on things like...banning TikTok. Despite the fact that we don't have a national data privacy policy.**

**What's interesting is that we *do* have laws on the books in the campaign space that regulate which entities can do what with people's data. So, in the future, if we do wanna try and pass a data privacy laws, maybe we can look to campaign finance laws as inspiration.**

**And according to Tara McGowan, real change is possible. It's just that clicking "Accept" and moving on with our lives has just kind of been the vibe lately.**

**TARA MCGOWAN:** You know all of us are blindly accepting all cookies on every website and letting people follow us and track us without thinking about the implications. I just don't think anybody's thinking about it and they're not necessarily feeling the negative impacts because there are positive impacts. There's definitely a feeling of helplessness towards it, which also just leads to numbness, which is not a good thing. It's a problem.

**RAFFI VO: But the thing about the "vibe" is... you never know when there might be a shift.**

**RAFFI VO: Next week on Technically Optimistic, we're talking about how data is used in health care.**

**Electronic health records means that healthcare is a data privacy issue for all of us.**

**But we're gonna spend some time talking about *reproductive* health care in America. Especially after the 2022 Supreme Court decision that overturned Roe v. Wade, and changed everything.**

**That's next time on Technically Optimistic.**

**[CREDITS]**

**CREDITS:**

**Technically Optimistic is produced by Emerson Collective. Production assistance from Christine Muhlke. With original music by Mattie Safer. Our senior producer is Erik Geannikis.**

**If you've enjoyed the podcast, please rate and review us – and subscribe on Apple, Spotify, or wherever you get your podcasts. Follow along on social, @emersoncollective.**

**And sign up for the Technically Optimistic newsletter! You'll get my thoughts about the week in tech, with lots of big questions, interesting links, and, you know, tons of ways to get in on the conversation. Subscribe for free at [technicallyoptimistic.substack.com](https://technicallyoptimistic.substack.com).**