Rationally Speaking #182: Spencer Greenberg on “How online research can be faster, better, and more useful”

Julia Galef: Welcome to Rationally Speaking, the podcast where we explore the borderlands between reason and nonsense. I'm your host Julia Galef, and with me is today's guest, my very good friend Spencer Greenberg.

Spencer is a mathematician with a PhD in Applied Math from NYU, where his work mainly focused on machine learning. Now he is running a startup incubator, called Spark Wave, that focuses on projects with significant positive social impact.

What that means primarily is: ways of doing social science research that are faster, and more rigorous, and more directly socially useful than is the norm. His projects are ... The main project at Spark Wave, through which he runs this research and publishes modules using the research, is Clearer Thinking. You can check out his work at clearerthinking.org.

That's what we're going to be talking about today, ways to improve the way social science research is done and make it more useful to society using the internet and/or machine learning. Spencer, welcome to the show.

Spencer Greenberg: Hi Julia, thanks for having me.

Julia Galef: It's a pleasure. Why don't you start with some examples of the kinds of projects you're working on at Spark Wave. Specifically at Clearer Thinking.

Spencer Greenberg: Sure. Very happy to do that. What we do at Spark Wave is we build technology to try to do better, faster, more rigorous social science research that can help impact people. We also actually carry out that research and then build tools from it.

To give a few examples, one of our projects at Clearer Thinking is, we're beginning to study habit formation, so we're looking at how can we get people to actually stick to new behaviors, which is obviously incredibly important.

We also look at decision making, so we actually built a tool that we're going to release soon on our website clearerthinking.org, that will help walk you through a major life decision like, "Should I marry my partner,” or “should I quit my job," and tries to help you do a better job making that decision so that you can get the things that you value.

We also do some more pure research. Like we did a bunch of research on Trump versus Clinton supporters and what variables best predict their support. We're doing research on overconfidence. We also productize our research to try to actually directly build products out of it that can be beneficial.

One of our big projects at Spark Wave is a tool to help people with depression. We've done a ton of research on how do you automate depression treatment, and can you actually provide a software that can benefit a wide range of people.

Julia Galef: Cool. Do you have any results that you can share yet from some of these studies?

Spencer Greenberg: We had some really exciting initial results on our depression app, which is called Uplift. We actually -- and this is just a pilot study so far, but in our pilot study we were able to reduce depression symptoms by about 50% on average over 32 days. So that was super exciting.

And we actually did a follow up, where six weeks later, with the same people, and we found that their depression symptoms were at basically the same level that they were at at the end of using our tool. So that was really exciting.

Julia Galef: Before you go on, roughly how many people in the depression study?

Spencer Greenberg: So, the numbers I was just giving you was for the first 80 people that completed our entire program.

Julia Galef: Got it. So for those 80 people, the average reduction in ... How are you measuring depression? Is it like a self report, on a scale?

Spencer Greenberg: We use a very common way of measuring depression, it's called the PHQ9, and it looks at nine different symptoms and you get a score on each of the symptoms, and then average those together. And so that's what we're trying to reduce. We're trying to reduce depression symptoms.

And so a typical population would have maybe a three to four on the PHQ9. We recruited a bunch of people with mild to moderate depression, so they started at like a nine, and then they went down, on average ... Maybe they started at nine and a half, went down an average of something close to 50%. So we've got them actually pretty close to what you'd find in the general population.

Julia Galef: That is a large effect size, for social science, as I'm sure you know. That's, most effect size are not ... Like even if they're statistically significant, they're not that practically significant.

Spencer Greenberg: Well, I have to say, I almost fell out of my chair when I saw the results. Obviously, we believed that it was going to work, or we wouldn't have invested so much time and energy in to it, but yeah. It was really exciting.

Julia Galef: What kind of interventions were you doing to reduce depressive symptoms?

Spencer Greenberg: So what we're doing is automated cognitive behavioral therapy. So that means -- it's not therapy like you do at a therapist, but it's taking the principles and

ideas of cognitive behavioral therapy, which is actually the most evidence based form of therapy for depression and anxiety, and what we did is we extensively studied the evidence on it -- looking at what's known about which parts of it work, why it works, what's limited about it -- and really tried to build a program that kind of walks you through that evidence based intervention, but with pure software.

So it's constantly adapting to whatever you say. So for example, it gives you homework assignments. If you didn't do your homework, it tries to help you figure out why you didn't do it, it tries to help you develop a strategy, so you'll do it next time. Did you forget, and therefore you need extra reminders? Or maybe you didn't understand the homework, so you need a better explanation, etc.

Julia Galef: The homework you're talking about here is part of the therapy? Like assignments people are supposed to do that will help with depression?

Spencer Greenberg: Yeah, it's part of the ... Yeah. Cognitive behavioral therapy is very, very different from most kinds of therapy, because first of all, it's very goal directed. You have a specific purpose that you're doing it for. Second of all, it's very evidence focused.

And third, it's about turning you into the therapist, in a sense. In other words, it's about empowering you with skills, that you can then apply to improve your life. Rather than ... If you think about it, if you go to a therapists office, you're with them, let's say, an hour or even less a week, versus how much time you spend with yourself, dealing with your problems. If you are taught to use these skills, than you can apply them every day, rather than just once a week. And it's so much more impactful.

Julia Galef: So I don't know if this is like a fair or coherent question, but do you have a theory of what counts as positive social impact? Is there any kind of framework you're using when you decide like, "This is a project that ..." or "This is software, this is an intervention that, if I could get it to work and be scalable, would have really large, positive, social impact." Or is it just a, "I know it when I see it," type question?

Spencer Greenberg: Well, I think there's direct suffering reduction, which is something I care a lot about, right? I know how horrible it can be to have depression, and I'm extremely motivated to eradicate it. So that's like the really clear cut case. Some of our work is less clear cut. Like another thing that, personally, I care about a great deal, is helping people make really good decisions. And so that's why we built this tool for decision making and we've done research on decision making. Now, it's not as directly clear cut. It's not like you're literally making someone suffer less immediately, but the idea is that in the long term, making better decisions will have this ripple effect in two ways. One, it's on your own life; you make better decisions and you're happier and you live more the life you want,

but second, on society where, in the world we live in has been growing increasingly complex and difficult to deal with, and I think people really, as a society, we need to make better decisions. We need to be better decision makers, when we make these really hard decision that effect thousands or million or in some cases, billions of people.

Julia Galef: Have you measured the impact of your decision making tool yet? Or is that still preliminary?

Spencer Greenberg: Well, so we've actually been teaming up… to do some really interesting studies, where we'll actually see if people use our decision making tool, have better decision making performance, in the sense of are they happier with the decision they made? They feel that it was better. Compared to a prior decision of theirs that was similar. Because of course, it's tough to have someone who comes and is making two decisions and we randomize and use our tool on one and not on the other. That's tough to do as an experiment, but we can compare it to a previous decision that was similar. So that's actually what we're gearing up to do. But actually, we also do research kind of internally, in order to figure out the details of our programs. I really kind of, an example of that that I was really excited about was, there's this problem where, when people are trying to make a decision, they often don't consider enough different options. So they kind of narrowly frame their problem as, "Okay, I can either quit my job or stay at my job. What should I do?"

Julia Galef: Right.

Spencer Greenberg: We knew that this was a problem, and so what we did is, we ran a little randomized control trial, where we randomized people in to two conditions. In the first condition we said, "Hey, there's this big problem where people don't consider enough options. We're going to ask you to wait right now, we're actually not going to let you continue, we're going to put like a timer, and ask you to spend the next 30 seconds generating some more options." The other group, we were even meaner to them. We said, "Hey, there's this problem called narrow framing. We literally won't let you continue, unless you generate at least one other option, for things you could do. And don't worry about being amazing, just you have to generate at least one." So can you guess what happened, when we ran this experiment?

Julia Galef: So you're comparing the people who were told to wait ... Was there a control group, who just said their-

Spencer Greenberg: No, in this case ... Well, we had previously run a study where we didn't have either of them, so that was sort of a control group, but in this case you were either asked, you're like, "You can't proceed for 30 seconds. We ask you to use that time thinking of other options," or we literally force you to, and we won't let you continue unless you generate on. Can you guess-

Julia Galef: I guess, I guess I would expect the latter group, the group that had to come up with an option, to be more likely to generate an option that they liked, relative to their previous, like the initial options that they had.

Spencer Greenberg: Yeah, good intuition there.

Julia Galef: Well, just 'cause it's too easy to write gibberish if you know that's all you have to do... Or just to not write anything, if you know that you get to continue in 30 seconds. Or maybe even to quit, I don't know, if you're forced to wait.

Spencer Greenberg: Yeah, so it turned out that virtually none of the people in the wait group, where they had to sit and wait, actually generated any options, but when we forced them to, every single person generated an option, and 25% of them ended up choosing that option, at the end of the program, instead of one of their originals. One in four people chose a different choice, because we literally forced them to make another option. Which I think is just mind boggling,

Julia Galef: Yeah. Do you have any shareable examples of alternate choices that people generated, when you forced them to?

Spencer Greenberg: Oh, that's a great question. I wish I had the data.

Julia Galef: Could just be a representative example, it doesn't have to be a real example.

Spencer Greenberg: Yeah, so unfortunately I don't it in front of me, but let me see if I, I don't know if I actually remember any of them…

Julia Galef: Like, in the job case for example. If someone was trying to decide, they were like, "Okay, should I quit my job or stay at my job?"

Spencer Greenberg: Oh, so I actually totally have an example of this.

Julia Galef: Okay, great.

Spencer Greenberg: So, funnily enough, I was reading the data recently and there was one person in there who, I think that their choice, if I remember correctly, was like, "Quit my job or leave my job." And basically what they-

Julia Galef: Wait, you mean, "Stay at my job?"

Spencer Greenberg: I'm sorry. Sorry. "Quit my job or stay at my job."And what they ended up realizing is that they could basically renegotiate their current position, instead of doing one or the other, kind of extreme choices.

Which to me, which is totally reasonable and you'd think they would have thought of that initially, but if they'd been in that wait group, they probably would have never though of it, probably, but because we forced him to think of

another, it came to mind. So that's a kind of way the online research I think can be so powerful, is that you can do these kind of microrandomizations. It doesn't have to be some massive study that takes weeks or months, and then you spend another two months writing it up as a paper. It could just be, "Let's study this one aspect of what we're doing, and see if we can get a better answer than we have before."

Julia Galef: Okay, so you're running all of these studies online, you're turning them in to actual tools. How far can this process actually go?

It seems to me that a lot of social science research involves people coming to a physical location, and experimenters observing the way they behave... This is a stereotypical social science experiment type thing. You have a bunch of people in a room and someone is, what's it called, they're like a plant, basically, by the experimenters. They're the one who claims that, "Oh, those two lines look the same length to me." And you see if that blatant lie, that is clearly contradicted by what we can see with our own eyes, causes other people in the room to say, like, "Yeah, I guess the lines are the same length." That kind of thing.

Or, you know, researchers observe, "Oh, when people are given a hot cup of coffee to hold, does that change their moral decisionmaking?" That kind of thing.

So the stuff you've been talking about seems like examples of things where, people are being given suggestions for things they can do, or they're being asked questions to reflect on, but how far does that go? How much of social science research could actually fit in your framework?

Spencer Greenberg: Well, that's a great question. So, there's different answers to that, and I think there's kind of interesting nuance there.

First of all, you could ask, well, of the current experiments being done, how many of them could you do online? And the answer is quite a few, but there's definitely ones, like the examples you gave, that you just couldn't do that experiment online.

But there's a deeper thing going on here, which is that the purpose of those studies is not to prove that holding a cup of hot water causes XYZ effect, right? It's trying to test some other high level concept. So there's a deeper question, which is, could you design an online experiment to test the thing you're trying to test? Right? Even if the current way of testing it isn't online. And so I think there's a much broader class of things you could test, with a cleverly designed online experiment, than are currently being. Right?

But the other thing I would say is it's sort of the issue of, if you lose your keys in the dark, but there's a lamp in one corner of the room, look under the lamp first. Not because your keys are more likely to be there, but because if they are,

it's going to be a heck of a lot easier to find them. That's how I view online research as well, is that with online research, we can do it incredibly quickly and cheaply.

And we can do this constant iteration, where we design one study, run it, modify it, run another, run another, run another, trying to converge towards the truth. Because we can do it so fast, rather than having to put so many resources into doing one that we can only do one every three months, right?

So basically, it's a very powerful flashlight, to peer in to human psychology and how to help humans be happier and make better decisions, that kind of thing. So I think that's part of the reason I'm really excited about it, and the last thing I'll just say about it, is that I'm really excited about online research, because it can directly turn in to apps and online tools, which can be deployed effectively for free. It's software. So if you find something, you don't have to go teach it to a person face to face, or somehow try and convince people to apply it in their own life, you could actually give them a tool that helps them do it.

Julia Galef: Right.

Spencer Greenberg: So it blurs this distinction between research and application, right? If we design a study that causes a certain affect, we could literally deploy that now as a tool.

Julia Galef: Right. Are you mostly using Mechanical Turk to test your tools, or to test the interventions?

Spencer Greenberg: Well, we do love Mechanical Turk. It's a wonderful source of participants for studies. We also, Clearer Thinking, we have a mailing list, with about 23,000 people, so that's a really good resource as well, because we can sort of test it on our own people who are interested in our products and want to try them out and give us early feedback and that kind of thing. So that's also a really great resource.

And then we actually built our own technology, to aid in research. So we have a tool called Task Recruiter, if you want to check it out, taskrecruiter.com, and it's basically a research tool that lets you recruit people to do studies. So we use that extensively in our own work.

Julia Galef: To what extent are you worried about the population of people, like on your mailing list or the population of people on Mechanical Turk, not being representative of-

Spencer Greenberg: I love that you asked that question.

Julia Galef: Of course I ask that question. Well, it's like Mechanical Turkers are, they're people for whom it makes financial sense to earn a small wage, answering

questions online. And they're Americans, mostly, I think. Actually, I'm not sure, but probably they're not representative of the general population.

Spencer Greenberg: The U.S. and India, are the two populations, yeah.

Julia Galef: Yeah, okay. That makes sense. Yeah. And now, I should be clear that, when you read social science studies in a typical journal, that population, or the population who participated in those studies is also not representative. They're probably less representative than M Turkers.

Spencer Greenberg: 18 to 22 year olds college students are less representative. Yeah. No, I totally agree.

Julia Galef: But your mailing list is probably over representing higher socio economic brackets and the U.S., but it's also selecting for people who are interested in decision making, and that's probably gonna have certain cognitive components, or correlates as well. So anyway, I'm just wondering what you think about the whole representative population issue.

Spencer Greenberg: I think it's a great question, because this is something that plagues all of social science research, which is that ... So let's say you go and you show that doing this particular thing to these particular people in this particular setting in this particular way at this particular strength, causes some effect, right? And then someone else tries it in a slightly different population or a slightly different way of actually doing the intervention, or whatever and they don't find it that good.

So what do you really conclude from that? Do you conclude that the original paper was a false positive? Do you conclude that actually, it only works on some subset of people or it has to be done at a certain strength? It's a huge problem.

We actually have a way of cutting through that problem that I find very appealing. Which is that we like to test our tools on the people we actually want to deliver the tools to. And so if you can get an audience that's really similar to the people you're trying to reach, then you can solve the generalizability problem, right? So that's what we do.

So testing on our Clearer Thinking audience is fantastic for us, because those are the number one prime consumers of our tools. Of course, we're always trying to grow and reach a broader audience, and there are lots of users of our tools that aren't on our mailing list, but it's pretty darn representative of the people we're reaching, right? So we are building tools to help those people.

And Mechanical Turk, similarly, so Mechanical Turk skews a little bit younger than the general population, more tech savvy than the general population. There's kind of these known skews it has, but what's cool about that is actually, a bunch of the skews are similar to our audience we're trying to reach. So, it's certainly way more representative than college students would be for us. It's

not perfectly representative, that's why we like to do a combination of testing on Mechanical Turk, also testing on our beta tester list, which are like our particularly excited active users. And so that's how we actually target the group we really want to reach.

Julia Galef: I was going to say, "Well, it sounds like you are engaged in a different project from academics who are looking at similar questions about depression or decision making, etc."

Spencer Greenberg: Oh, we are. We're definitely engaged in a different project.

Julia Galef: Well, I mean, okay. The way in which, the kind of difference that I was going to bring up, is that academics are, at least in theory, trying to derive these causal models of human psychology. Like, "This is how humans behave under these situations and this is why. This is our explanation for what's going on, that makes them behave this way." And that might have implications, but the main goal there is explanation. Whereas it sounds like your main goal is useful interventions, ways to reliably intervene on human psychology, to cause better results. Is that an accurate ... I've been going back and forth as I've been speaking this paragraph, as to whether that's a good characterization or not. What do you think?

Spencer Greenberg: Yeah, I think that's right on. Basically, if you start with the goal of helping people make better decision, helping people be happier, suffer less, and then you work backwards, you get something that doesn't look the same as what academics are doing. That's not the process that they are going through. Obviously, some of the are motivated by helping people, but the incentive structure in academia pushes in a certain direction, which is towards generating new insights, publishing papers in top journals, and if you're not doing that, than you kind of get kicked out of academia. SO you don't have much choice.

Julia Galef: But, I mean here's why I was hesitating. I framed that as if you're doing one thing, they're doing the other thing, but doesn't your thing also do the explanatory thing? In what way are you not producing explanatory models of human psychology?

Spencer Greenberg: Well, it's not the prime thing that we focus on. It sometimes is useful and we do try to do it when we need it, but for example, let's suppose we try something and we find it works. We don't need to take the next step and actually figure out why it works. If knowing it works is sufficient and we can plug it in to our tool and we can prove that it has the effect that we want, then it's gravy if we can prove why, but it's not our fundamental goal. Our goal is to make it helpful, trying to apply it in an effective way.

And sometimes we do try to understand phenomena, because we think it kind of aids what we're trying to do. And we'll look at the academic research on topics and look at the theories they have. A lot of times we find, when we do

that, is we get a lot of good ideas from academia, but usually, at some point, we kind of hit a wall. We're like, "Okay, well we've looked at these different theories, looked at these different ideas. It still doesn't tell us how to build this thing." It gives us some hints and suggestions, but there might be, for example, multiple theories that are not being coalesced and so you don't have a comprehensive theory.

That happens a lot. Or there's not enough detail, not enough specificity, that we can then actually build something. It's still helpful, but we have to go a step further. And actually, I have this idea I like to think about, of full stack social science. Have you heard the phrase full stack web developer?

Julia Galef: No, actually -- oh, I have heard of it. I don't know what it means.

Spencer Greenberg: So full stack web developer-

Julia Galef: This happens to me a lot in the Bay area. I hear these terms getting tossed around by my coder friends. I'm like, "Oh, I'm not going to bother interrupting conversation to make them explain it to the non coder." So what does it mean?

Spencer Greenberg: Sorry to throw out random jargon... So the metaphor doesn't work as well if you don't know what it is, but basically a full stack web developer is someone who works on both the front end, like the interface of an app and the way it looks and stuff, and the back end, so the databases and the kind of under pinnings of the app. Right?

And a full stack developer kind of goes between both of those, rather than being just front end or just back end. I like to think about this idea, I call full stack social science. Which is, okay, suppose you come up with great ideas, but you don't test them. Well, they're not going to be that useful, right? In social science.

Or, suppose you come up with great ideas and test them, but you don't actually go and do anything with them. Like you just say, "Okay, those ideas are out there." Is someone actually going to pick them up and apply them in a useful way? Maybe, if you're lucky, maybe someday, or probably not. Right?

And so you kind of have to do the full stack of activities in some cases, to really have the impact, in my opinion. In other words, you have to fill in gaps where nobody has just the right idea to solve that problem. You've got to do background research, just to understand the phenomenon better and test parts that maybe need to be tested -- but then that's not enough, you have to actually go build something out of it, and then you have to go market that thing and get it in the hands of people, so it can actually help them. So that's what I view as our mission, is the full stack of activities, so that we can change behavior.

Julia Galef: Well, I guess what I'm wondering now is, why wouldn't it just be better for academia to start with the question of "What are major problems that need to be solved in the world?" And then do research to figure out how to impact those problems, and then, from there, figure out why that's working?

Or do you think there's just additional unique value that's being added by starting with, "Let's build models of human psychology, and then later see how to apply them." It seems more indirect to me.

Spencer Greenberg: I think that would be a great thing for the world. I just think it's this different thing, though, than academia was ever doing or was ever designed to do. It's sort of a machine that tries to create understanding of the world, but I don't think it, as far as I know, has never been directed at specifically how do you solve problems in the world, which is a different thing.

And sometimes understanding does, later, eventually lead to some breakthrough. People classically quote how number theory was originally believed to be completely useless -- and actually, some mathematicians like that about number theory -- but then, oh no! They found this application in cryptography, now it's not a pure science anymore. Pure mathematics.

Julia Galef: Now they're not the cool ones at the math parties?

Spencer Greenberg: Yeah. It's just kind of different goals.

Julia Galef: Are you separating out your exploratory analysis from your hypothesis testing? I mean, a common critique of social science, as I'm sure you know, but for our listeners, is that if you just test a whole bunch of hypothesis at once, some of them are going to come out significant, even thought that's just noise, and so common exhortation is to first do a bunch of exploratory work, where you identify which hypothesis are worth testing, and then do sort of statistically rigorous testing, where you can say that if it comes out significant, then that is meaningful.

Spencer Greenberg: See, Julia, your question is making me so happy, as a mathematician. Yeah, so basically that is ... Multiple hypothesis testing is a massive problem in science, in general, and including social science. If you test too many hypotheses and you don't have enough data, you'll inevitably find something, that doesn't mean it's real. I take this extremely seriously. Our approach to solving this problem is we just keep running study after study, and we just make sure that we confirm things beyond any reasonable shadow of a doubt. When you're hell bent on figuring out about the truth of something, as opposed to saying, "I've got a deadline, I've got to get this thing out." Which is a luxury we have, where we can focus on the truth of the matter.

It just changes everything, because then, you're like, "Wait a minute. We think this is true, but we just came up with some possible, convoluted way that we

may have misinterpreted this. Let's go run yet another study." Right? There was one project where we ran 12 studies, because we kept finding ways where we're like, "Well, maybe we kind of didn't quite fully understand this." Right? And we haven't even published that work, it was just 12 studies to understand the phenomenon. It's a totally different mindset and this is part of why we build the technology behind running studies, because we know we have to make them so fast and easy for us to run, that it's not a big deal. We can run 12 studies, that's fine.

Julia Galef: So just to clarify, it's not just like you ran 12 studies, each one tested a slightly different version and you go with the one that came out significant. Something seems to work in the 11th study, then you're like, "Let's confirm it and run a 12th study," and if that one also comes out significant, then you're like, "Yes, this is real." Is that right?

Spencer Greenberg: Well yeah, in that case we were trying to pinpoint this phenomenon and we kept finding slippery ways where we were like, "Well, but what if this? What if we misinterpreted this? Or what if this is just happening because of this other phenomenon?" So we kept trying to find ... Basically, we kept coming up with ways maybe our original conclusion was wrong, and trying to find studies that would test if we were actually mistaken. So that was basically the process of trying to hammer that out. It gook so long. Yeah.

Julia Galef: I was just going to say, I don't think it's widely appreciated that, a way to solve the p-hacking problem, is just to make it really fast and cheap to run studies.

Spencer Greenberg: Well, you know what? It's funny, 'cause it's, in a way it depends completely on intentionality.

Julia Galef: Whether you *want* to solve the p-hacking problem, you mean?

Spencer Greenberg: Yeah, exactly. That's what I'm saying. Because we are lucky enough and also motivated enough, just to be hell bent on figuring out what's true -- for us running lots of studies is this wonderful thing, 'cause it helps us poke holes in our own work and get a deeper understanding of reality.

But suppose that your goal was to get a paper out by a certain date, right. Then running more studies is a way to just find something that works, that looks cool. Right?

Julia Galef: Right.

Spencer Greenberg: It completely depends on, it can be used for good or evil, right?

Julia Galef: Right.

Spencer Greenberg: And there is a problem, where sometimes people will run a bunch of studies and only publish one of them, but the other ones kind of semi contradict it, the one that ended up published. Not necessarily in a fraudulent way. Usually, it's not fraudulent. Usually it's just kind of, there's a rationalization by which maybe it's okay, but maybe it's not okay, right?

Julia Galef: Yeah. When you were talking about noticing little ways the study might be wrong or might be measuring something other than what you wanted, I remember you telling me something a while back about asking people how they were interpreting your questions and continuously getting struck by the fact that people were interpreting your questions very differently than you thought. Is that what you were talking about?

Spencer Greenberg: Well, yeah, that's one example, and one that's very dear to my heart, because -- okay, I'll give you an example. We were trying to replicate, it wasn't an exact replication of [the literature], but we were trying to do something similar. Where we were looking at the sunk cost fallacy, and we were trying to design questions that would elicit the sunk cost fallacy. For example, on our website-

Julia Galef: Can you just explain what the sunk cost fallacy is?

Spencer Greenberg: Oh yeah, absolutely. Yeah. I forgot that that's not necessarily a thing-

Julia Galef: No, it's fine. I bet a lot of my audience knows this, but for those who don’t.

Spencer Greenberg: Yeah, so the sunk cost fallacy is basically, it occurs when you invested a lot of time or resources or money or just emotional energy in one particular path. For example, you're in law school and you've already spent the first year of money on law school and you're thinking of yourself as a lawyer, and then you realize that actually that path is no longer sufficiently valuable going forward, to make it worth it.

Julia Galef: Like the job market's so bad, you don't actually expect to get a good job. Or you don't expect to enjoy it or something.

Spencer Greenberg: Exactly, but then the problem is that if you start to think about leaving law school, you start to have to register all the investment in law school as a loss. Whereas if you just keep on with the same path, you could pretend like you haven't lost that value you invested.

Julia Galef: Right.

Spencer Greenberg: Of course, it doesn't make any sense, from a rational perspective, because in this case we stipulated that you're better off leaving law school. It's just about how you process the past loss, and it's gone either way, regardless of what you do.

Julia Galef: Right.

Spencer Greenberg: So the sunk cost fallacy, at clearerthinking.org, which we have a program that trains people to help avoid the sunk cost fallacy, and so we did some research on it, and one thing we were doing is we were trying to develop a question that would elicit the sunk cost fallacy.

Where we would ask you, Okay, suppose you're at dinner and you order some food and you realize though that you're actually not at all hungry. You're totally full and you cannot reasonably bring the food with you. Would you just finish eating it anyway?

And we thought, "Oh, maybe this will elicit it," because people will view it as, "Oh, if I don't eat the food, then I've wasted the food and-"

Julia Galef: And the money I spent on the food, yeah.

Spencer Greenberg: "And the money I spent on it," and maybe that would elicit the sunk cost fallacy.

So it turned out, it caught tons of people -- "Oh look, these people are irrational. Sum cost fallacy." But we were lucky enough to also ask those people to explain their answer. Like, why would you say that you would keep eating this food?

Julia Galef: I don't think I want to call that lucky. I think that was a good, that was a smart choice.

Spencer Greenberg: Well, *now* we always do that. Now-

Julia Galef: Okay, I see. So the time-

Spencer Greenberg: We've learned. We've learned that you need to always study your own questions, to try to understand. Not just look at someones answers, but understand why they're answering that way. So we love asking people why they gave an answer they did, to better understand.

And so, in this case, all these people that we thought were being irrational, can you guess what they said? Why they would force themselves to eat this food, that they didn't like the taste of, when they were already full?

Julia Galef: Hm, maybe they said that that they were trying to teach themselves a lesson or something? That this would make them less likely to just sort of blithely order large amounts of food the next time? That seems kind of masochistic, but-

Spencer Greenberg: No, that's an interesting explanation, and I bet some people do think that way. But by and large, the most common reason people said, is because they assumed they'd be eating with someone else. And they thought it would be

awkward if they didn't eat their food. It had never even occurred to us they would assume that, right?

Julia Galef: That totally makes sense. I didn't ... Yeah. That makes a lot of sense.

Spencer Greenberg: So then we changed the question. We stipulated that they were not eating with someone, that they were alone. And then, a handful of people said-

Julia Galef: Can I guess?

Spencer Greenberg: Yeah, yeah.

Julia Galef: Did they not want to make the waiter feel bad?

Spencer Greenberg: Very close. They said, usually they said the chef.

Yeah. This is where it really hammered home for us, that unless you had ... A multiple choice question can be a very dangerous thing, because it gives you the sense that you've measured something, you get a score, but if you don't understand the *why* behind it, like, "But why are people putting six instead of three?" Right?

Julia Galef: Right.

Spencer Greenberg: Then maybe you're actually really misleading yourself.

And this actually segues me to another example of this, where we were -- in this case we were directly trying to replicate something in academic literature. It was, someone had developed a scale, saying that way more people than you'd expect had these kind of delusional beliefs, of sorts.

And we were really excited about this, because we thought hey, maybe this would present an opportunity to help people, if people really do have these delusions. Or it's also just an interesting and important thing to know about.

So we ran a study and it turned out that we were able to replicate a bunch of their findings, in the sense that people did report these delusional beliefs. For example, people reported that they feel like bugs are crawling all over their skin, and other kind of strange things like this.

But we also asked them *why* they answered that way. And well, it turns out a bunch of people have lice.

Julia Galef: Oh. Oh my God. That's horrifying, and also kind of funny. Wait, a *bunch* of people? Really?

Spencer Greenberg: Well, they weren't saying that 50% of people had this, but they were saying five or 10% of the population-

Julia Galef: Have got it?

Spencer Greenberg: Yeah, they were saying, more than you'd expect. Right? You wouldn't expect 10% of people to say that ... Or sorry, 5% of people to say that they have these delusional bugs crawling over their skin, but 5% of these people might have a bug infection-

Julia Galef: Or maybe bed bugs.

Spencer Greenberg: Or bed bugs, exactly. But what we realized is that almost every one of these things, there was actually a fairly reasonable explanation. At first glance, it seemed like these people must be delusional. It was a huge lesson for us.

Julia Galef: I know I'm asking you to sort of speculate wildly here, but do you have any guess as to what percentage of results, significant results reported in social science literature, would be significantly undermined if the researchers just asked subjects, "Why did you answer that way?"

Spencer Greenberg: Yeah, I really have no idea, but I'll give you an interesting anecdote about that. You know the endowment effect, where people ... Basically, if you give someone a mug and then you say, "How much would you sell this for?" versus if you show someone the mug and you say, "How much would you buy this for?" they give really different numbers. It's called the endowment effect, it's viewed as a similar idea to cognitive bias of some sort.

Julia Galef: You mean like, you would have to pay them a lot more to give up the mug than they would be willing to pay to get it?

Spencer Greenberg: Yeah. Like the fact that you were just given it, suddenly makes your value of it increase dramatically.

Julia Galef: Right.

Spencer Greenberg: For some reason, right? And this has been kind of well studied.

And just recently, some researchers, this wasn't the only piece of evidence they had, but one of the things they did, was they asked people why.

Julia Galef: Okay.

Spencer Greenberg: And they found that a bunch of people were saying, "Well, I don't want to get ripped off. I feel like that I have this mug now, I don't want to feel like I was cheated by being paid too little for it." Which is not the same thing as the original theory behind the endowment effect, which is more around somehow

we intrinsically value the things we have versus the things we don't have, or something like that.

Julia Galef: I mean, it's still like, I don't know if I want to call it irrationality, but it's still not a pure “trying to maximize the value of my possessions.”

Spencer Greenberg: But think about this. Let's say they're aware, vaguely aware of a market price, right? If you say that, "Oh, the market price for this mug is X, I don't want to sell it for less than that." It's just sort of a different explanation, right?

I guess whether it's rational or not, is sort of, is an interesting point, but it's sort of secondary to the point I'm trying to make, which is just that, they asked the people and the people gave a different explanation.

And there's actually other evidence now that that might be a more accurate explanation too. So who knows. I don't know. I don't know if it deals with the endowment effect, but it's pretty interesting.

Julia Galef: But I guess one way to estimate this is just, how often it turns out when you run studies and you ask people why they answered the way they did, that it turns out that the way they interpreted the question, or the context in which they were thinking about it or whatever, was different than you assumed when you were running the study. If that is a common thing that happens for you, I can assume, probably, it would be common for other researchers *if* they tried to check.

Spencer Greenberg: Yeah, I would estimate probably maybe about a third of the questions we asked, we learned something. It doesn't necessarily mean we ended up changing the question, but we kind of learned something maybe that we didn't know.

Julia Galef: It would change your interpretation of the results?

Spencer Greenberg: Yeah, well maybe a little bit. But there are, sometimes it's just, we just actually literally just get rid of the question, because we're like, "Yeah, this is not at all ... We thought we were measuring X, we're actually measuring Y."

And actually, this leads in to a topic, which is quantitative versus qualitative research, which we didn't talk about really. In my opinion, very often these things are separated. Some people are really into quantitative research, other people are really into qualitative research, where they like doing a lot of interviews and things like that.

I think they're both so valuable and useful, and they work so well in synthesis. They kind of really enhance each other, and I think it's a shame that they're not run in parallel more often.

And I think this is a good example of that. Because asking people “why,” is actually qualitative data, right? You're getting free-form text responses, you have to literally read and interpret. Whereas multiple choice questions, they're quantitative data. You could run an algorithm over, right?

Julia Galef: Right.

Spencer Greenberg: And so, we really like this back and forth. We'll design a question, ask people to answer it, but also ask them why -- and then use that to redesign the question and then get a measurement and kind of so on.

Julia Galef: Cool. Yeah, okay so we've talked about a number of ways that running studies online can help make research faster and thereby better, or just directly make research better. Another big category that we haven't really talked about, is making research more *open*, in the sense of open science. We had Brian Nosek, from the Center For Open Science, on recently. We talked about-

Spencer Greenberg: Oh, Brian's great. Yeah.

Julia Galef: Yeah. You, I believe, are working on a way to systematically make research more open, for researchers, using your online study tools. Can you talk a little about that?

Spencer Greenberg: Sure. When it comes to open science, where you're basically making it clear to the world what you did and what your results are, and also making it easy for other people to replicate those results. Also making it easy for people to get new insights from those results. One of the challenges is that a paper just doesn't contain enough information, in a certain sense.

Julia Galef: Right, right.

Spencer Greenberg: And part of it is that sometimes journals will just limit you. They'll say, "This is how many pages you have." And you're like, "Well, but I did all this great research," and you have to cram it in, right? In other cases, it's just because you don't want to distract from the main focus.

Julia Galef: Although in theory, you could have online appendices. Like, I feel like the excuse that there isn't enough space in the journal is not really ... If researchers actually cared about being transparent, there are relatively easy ways to do that, and they don't.

Spencer Greenberg: Yeah. I totally agree, but think about the incentives -- unfortunately they don't get any benefit, right? They can go put it up on their personal website, but if the journal doesn't ask them for it, or make them do it, and there's not even a place the journal offers to put it… they can put in on their blog, but will anyone see it? And is it worth that extra time if they don't get any benefit? Yeah.

Julia Galef: Brian's working on that, but there's a long way to go.

Spencer Greenberg: Exactly. Yeah, Brian is definitely working on that.

So, one thing that we think about, one of our tools that we're building, it's called Hypothesize. It's to help people do statistics.

Our idea is that, to a shocking degree, people often make errors when doing statistical analysis, even people that you'd think wouldn't make errors. And this is because statistics is really complicated and is not really well suited for the human brain -- and also because a lot of people work with statistics, they weren't originally trained as statisticians, right? Ironically, a lot of times statisticians don't deal with data at all, they just kind of theorize, and the people dealing with data are not statisticians.

So anyway, so one thing I really like about the idea of building tools to help people with statistics, is that then, the statistics you do could be a permanent link basically. So I've completed a statistic, an analysis for my paper, I could literally have it be a link. I could put that link in my paper, someone could click on it, go see every statistic run. Not just the few that I put in the paper, but see the full gamut of them, and potentially even change the assumptions and see how that would have updated the results.

Julia Galef: Oh, to see how robust they are. Or not just robust, but how context-dependent they are, what context they apply in, etc.

Spencer Greenberg: Exactly, or even just maybe they're interested in a slightly different analysis, that could help your own work. And so that's just one -- this is a small piece of the puzzle, but I think that it's one potentially helpful thing that could help people better grasp what's happening.

And then, related to that, we also built a platform called Guided Track, that helps people build on my experiments. The idea is to try to give people a super powerful tool, to build kind of completely automated experiments and interventions, and we use it ourselves heavily.

But one of the neat things about that, is that once you have an automated intervention that you built, you can then put a link to that in your paper, and someone could -- we actually have an option, you can set it to public -- so someone can go to your experiment, they can view it from exactly the experience that the participants got, which is really important if you're trying to replicate work. They can also look at the code of it, if you made it public, and there's a button to copy it.

So now you can just say, "Oh, I want that experiment. Copy." You go make your own tweaks and then you run your own version of it, that now has some different variables.

Julia Galef: So let's say you just wanted to try to replicate it directly, and not make any tweaks, could you literally just click a button and replicate it?

Spencer Greenberg: Well, this is part of our long term plan, but we're not there yet. We're not there yet. Yeah, I mean basically, if you take our tool like Task Recruiter for recruiting people online, you take Guided Track for building online experiments, you take Hypothesize, our statistics system… Yeah, this is what we're converging towards.

Ultimately we want one click replication. You click a button and you can completely re-replicate an experiment, recruit new participants that match the demographics, that are supposed to be recruited. Automatically get them in to the experiment. The experiment automatically runs on them. All the data's collected. Automatically statistically analyzed, based on the original system analysis you chose. Reports generated -- and yeah. There you go. It's like Amazon one click, but for science.

Julia Galef: That's very cool, but making open science easier to do is sort of only one side of the problem. There's also the problem of giving people the incentive to make their work easy for people to try and replicate, which as you were pointing out earlier, is not really built into the incentive structure of academia.

This is also something Brian and I talked about on his episode. People get rewarded for publishing significant results and publishing a lot of them. That's how they get approval and citations and tenure and so on and so forth. Is there any way that your tool addresses the incentive side of the problem?

Spencer Greenberg: Well, I'm a big fan of nudges. Basically, it's very hard to force anyone to do anything, but you can give them nudges in the right direction, and sometimes nudges can do a huge amount. There's interesting research on organ donation, how flipping it from defaulting to an organ donor to defaulting not being an organ donor can have a huge impact, it seems.

So if you take the idea of, okay, well right now, it's all this extra effort for someone to publish their data, or publish their statistics, or publish their codes and make it easy for someone else to take that experiment and copy it. It's so much effort to do these things, and the benefit is so little, if we can bring down the barrier of effort, really, really low, so that it's very easy… and in fact, if you're already using our tools, we can sort of automatically kind of nudge you to it, maybe even make you feel slightly guilty for not releasing your data to the world.

Julia Galef: Like, “click here to confirm that you are a bad actor who doesn't care about open science”!

Spencer Greenberg: No, no, but seriously. If you have someone’s data already in your tool, and you're like, "Hey, you know, by clicking this button, you could release this to the world so so many people could benefit from your data you've collected, we

could even put it on a time delay. Delay it for a year and automatically release it in a year, if you want to make sure you have time to write your paper." That's a very different scenario than if you're like asking someone to go do a whole bunch of extra work, or nobody's even asking them, right? Nobody's even asking them to do it at all, but they'd have to think of it on their own, and then do a bunch of extra work, instead of just us doing it for them automatically, right?

Julia Galef: Well, okay. I guess I was conflating a couple things there, when I talked about the incentive problem. There's the, “you get no reward for all this extra effort you have to put in” problem, but there's also the problem of -- I guess this doesn't quite fit under open science, but under good scientific methodology in general -- there's the problem that following good methodology and avoiding p-hacking and stuff like that, makes it harder to get significant results. And as mentioned, you get rewarded for publishing significant results.

So isn't giving people tools to make sure their statistics are all very rigorous, isn't that kind of counter to their interests?

Spencer Greenberg: Yeah, so that's a great question. So that's why it's so critical that if you provide someone tools that make their results more robust, those tools also have to make their research faster.

Julia Galef: To compensate.

Spencer Greenberg: To compensate.

Julia Galef: Yes.

Spencer Greenberg: Because more robust research means fewer publications, because you don't publish as much bullshit, right? As much fake stuff. Or you catch it before it goes out.

Julia Galef: Right.

Spencer Greenberg: More often. And so it has to be faster, in order to compensate for that, but I think it can also make it higher quality at the same time.

Julia Galef: Right, right.

Spencer Greenberg: You can trade these things off against each other, but I think that's really important.

Julia Galef: Right, so you end up, hopefully if all goes well, according to plan, you end up publishing significant results at a similar or even faster or greater rate than you were originally, but they're actually real, much more likely to be actually real. Or at the very least, much easier for other people to test if they're actually real.

Spencer Greenberg: Yeah, and you know another thing I'll say about that is I think there's really, if you're producing research that's shoddy, then you're building a foundation of a house out of sand, and then you're going to try to build upper layers over the next 10 years.

Julia Galef: Right, right.

Spencer Greenberg: The higher quality your research is, that actually costs more up front, but you're going to build a much better house, right? 10 years from now, you're going to build layer upon layer upon layer, and make a lot more progress and your later research is going to be much benefited.

And I think that's also a problem in society at large, when we have a situation where there's a lot of false positives. Researchers can't trust each other's work that well, and it's hard to build on what other people do, because maybe that's just not a good foundation to build on. And I think the whole process in general goes much slower, and it's no longer standing on the shoulders of giants. It's like you've got to check really carefully which giant you're standing on.

Julia Galef: Or if the giant has a gimpy leg.

…It's a messed up metaphor. Anyway. Cool. Well, we're actually rather significantly over time at this point, I just got caught up in this thread -- but before we close, Spencer, do you want to introduce the Rationally Speaking pick of the episode? It's a book or website or article or something that has impacted your thinking, in some significant way. What would your pick of the episode be?

Spencer Greenberg: So my pick would be the book *Feeling Good*, by David Burns.

Julia Galef: Ah yes, I remember when you were giving copies of *Feeling Good* out, to anyone who would take one.

Spencer Greenberg: Oh yeah, I have like 10 copies of it in my house right now. No, I think it's a wonderful book, that introduces people to the principles of cognitive behavioral therapy, which as I mentioned is the most evidence based therapy for treating depression and anxiety. It's not to say it's the only evidence based one, but it's the one with the largest quantity of evidence, showing it's effective.

Julia Galef: Right.

Spencer Greenberg: But what I think is really wonderful about it, is that it's about the, a lot of it is about the way you think, and you start ... So for me, one of the big changes it had on me is it made me much more aware of how when I'm upset, when I'm experiencing emotion, it changes the way I think about things. What would have been, I would have thought about something a certain way an hour ago, now I'm thinking about it a different way, and in maybe a much more harmful way.

I think CBT is a wonderful set of tools to help you peer in to your own mind and kind of find more helpful ways of thinking, and actually truer ways of thinking. And so I love that about it.

And that book, *Feeling Good*, is specifically focused on depression, so I especially recommend it if you're depressed, or if someone you know is depressed. But actually, David Burns also has a lovely book on anxiety, called *When Panic Attacks*. It's not just for people with panic attacks, it's for anxiety broadly, CBT applied to anxiety. If you have anxiety, I would highly recommend *When Panic Attacks*.

Julia Galef: Excellent. Yeah, you can have two picks, that's totally legit. I've had guests try to sneak in three picks as well, so you're still well within range. Cool. Excellent. Spencer, thanks so much for being on the show.

Spencer Greenberg: Thank you so much. This was great.

Julia Galef: We'll link to clearerthinking.org as well as to your two picks. This concludes another episode of Rationally Speaking. Join us next time for more explorations on the borderlands between reason and nonsense.