Rationally Speaking #180: David Roodman on “The Worm Wars”

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, David Roodman. David is a senior advisor for the Open Philanthropy Project. Before that, he served as the senior economic advisor at the Bill and Melinda Gates Foundation. He specializes in public policy and economics, especially economic development in poorer countries. David was a guest on Rationally Speaking previously about a year ago now, maybe a little more, when we were talking about microfinance and his book, *Due Diligence*.

This time, we're going to be talking primarily about something called the worm wars, which sounds most of all like a science fiction epic trilogy. But no, it's even more exciting than that. It's a controversy in social science.

So, David, welcome back to the show.

David Roodman: Great to be here.

Julia Galef: All right. Let's set the stage here.

Starting in about 2004, the economic development world got really excited about deworming pills as a cheap and easy way to help people in poor countries. The idea being that in a lot of countries, especially in Africa, it's quite common for people to be infected with intestinal worms, parasites, which interfere with their nutrition and cause illness and that in turn makes it hard for them to go to work and school and succeed in life.

But, we have pills that get rid of worms, and so the idea was that in a way this is like an anti-poverty pill, right? Not just can we give people a pill and make them healthier, but that that has ripple effects for their earning potential down the road.

The source of all this excitement is a study that came out in 2004 showing that deworming reduces poverty, which, to skip ahead in the story for a second, that paper was revisited a couple of years ago and its results called into question. And that controversy is what's been dubbed the worm wars.

But, before we get to the wars themselves, David, maybe you could kick things off by telling us about that original study that generated all of this excitement about deworming.

David Roodman: Okay.

That was a great summary. There are a few things I would tweak in there, but I could never could never have distilled it as nicely as you just did.

Julia Galef: Yes, please do.

David Roodman: The idea that intestinal worms are a bad thing and that it's worth spending money to get rid of them actually goes back more than a century. The Rockefeller Foundation, in its very early days, when it was sort of helping to define modern American philanthropy -- I guess there was also Carnegie and others doing that, too… But they became persuaded that one of the best ways they could use their money was to deworm kids, not in Kenya but especially in Southern United States. And they launched a multi-year campaign from around 1910 to 1915 to do just that.

Julia Galef: I didn't know that. Go on.

David Roodman: Yeah, this goes back far. But what happened in 2004 was publication of a paper in a top economics journal by Ted Miguel and Michael Kremer, where they had carried out a randomized study of a project to give kids deworming pills in Western Kenya. As in any randomized study like this, some kids got pills and some didn't. Or more precisely, some schools were visited for distribution of the pills. In other words, the way that they did this in order to keep the cost down was that workers would go to schools --

Julia Galef: Instead of peoples' homes.

David Roodman: Exactly, which would just take more time and cost more money.

Julia Galef: Right.

David Roodman: So you get kids who are school age and you just give the pills to everybody who's at the school. And actually it wasn't just the pills. There was also education about the importance of washing your hands and things like that to help change behaviors as well as biology.

Julia Galef: Actually, I'm not sure I know this stat, but what percentage of children tend to have intestinal worms?

David Roodman: It really varies a lot by location. In areas that we're concerned about, it can be as high as 20%, 40%. In the place where this study took place, I think the average number of worm infections -- because there are different kinds that you can have, so you can actually at any given time have more than one -- was I think about two, maybe two to three.

Julia Galef: On average. So some people had more than two types of worms in their belly. That's lovely.

David Roodman: This is in an area where the infections are really quite intensive.

Julia Galef: Yeah.

David Roodman: This study was actually most important not for what it told us about deworming but because it launched the movement to do randomized studies in development economics. It was the first big one, and I bet you've talked a lot about that in your podcast. Well, this is where it started.

Julia Galef: Yeah.

David Roodman: There have been randomized studies in other parts of economics, but they brought it to development economics.

Julia Galef: That's actually one thing I wanted to ask you. It's a bit of a tangent from deworming itself, but it's important.

Do you have any sense of why this transition happened when it did with this 2004 paper? I mean, there doesn't necessarily *have* to be some explanation -- Obviously, if a transition happens, it has to happen at some time. But, you know, it still seems useful to ask, "Why not earlier? Why not later?" Was it that no one in economic development before Miguel and Kremer had thought of doing a field experiment, like a randomized experiment, or was there something preventing previous researchers in the field from doing it?

David Roodman: That's a great question. Like a lot of questions about history, there are multiple answers. I don't feel like I've got --

Julia Galef: Sure. Yeah. Hey, this is a safe space. Feel free to speculate widely without any evidence!

David Roodman: There are a couple of factors that I can think of.

First, starting I think maybe in the 1970s, possibly earlier, there were some very important experiments run not in development economics in the United States and elsewhere within the field of economics. I guess there's a sense in which it's only a matter of time, if these were seen as effective, and they were, that they would take over all parts of economics.

I think something else that was going on at the same time is that around 1980 you had the arrival of faster and faster computers, and so there was this period in the '80s and '90s where a lot of the effort of empirical economists, as opposed to theoretical ones, went into developing fancier and fancier mathematical methods for trying to study patterns and data. I think then there was kind of a backlash against that, saying that all this fancy math was not actually working as well as it seemed and was mostly just confusing people and hiding the problems in the data. And if you really want to know whether x causes y, you gotta run an experiment.

And then the math is really simple, pretty much. You look at the average breast cancer rate in this group and you compare it to that group. And you just see if the averages are different.

So it was part of a backlash against the [theoretical] fanciness in economics. Actually, I think it's been a good trend.

Julia Galef: Yeah.

David Roodman: There's a sense in which economics is simpler today, but also more empirical and grounded in reality.

Julia Galef: Yeah, interesting. But then, before 2004 in economic development, how did people decide which things to try to reduce poverty in developing countries? Was it just common sense, "This seems like it should work"?

David Roodman: They would do non-randomized studies. A set of studies that I'm very familiar with were of the impacts of microcredit in Bangladesh.

Julia Galef: Right, yes, of course.

David Roodman: Yeah. That was based on data collected in 1992 and then a second round in I think in 1999, and there was no randomization there. It's just that in some villages there was more microcredit than in others.

Julia Galef: Right. So this wasn't sort of like pre-evidence based medicine when doctors were just relying on common sense or common wisdom. There were serious attempts to be empirical, just not in a randomized way.

David Roodman: Yes.

Julia Galef: Got it.

David Roodman: Also, there was the foundation being laid by people like Angus Deaton, who just got the Nobel Prize a couple of years ago, of methods just for measuring stuff we care about. Like what is called in economics consumption, which essentially means how much stuff you buy each month or each year.

In the west, we have all these statistics that we can track. If you want to know what the GDP was five years ago in the United States, you can get that. But in a developing country context, if you want to know how people are doing, there's no statistician around who's gonna hand you the statistics. So, years of work went into just developing methods for running surveys and then figuring what questions to ask and how to compile those answers into indexes to represent things like consumption.

Julia Galef: Yeah.

David Roodman: And so that had to happen before you could do any kind of research on what raises consumption long term.

Julia Galef: Right. That makes sense.

So we've talked about Miguel and Kremer's paper in terms of its significance for the field, but we should also describe what it found in terms of deworming.

David Roodman: Right. They looked at two outcomes, as far as I can recall. At least two major ones. There were probably others that were folded in.

One was whether kids came to school more after they had been received these deworming pills. Maybe I should explain some of the basics here.

These pills probably ... the recommendation is that you give them once, or maybe twice a year, more often in areas where there are more worms. They're seen as having essentially no side effects. Nothing serious, which means if you give the pills to kids who don't need them, not much harm is done. And, meanwhile, if you want to test kids to see if they've got worms before you actually give them the pills, you can spend more money doing that. That can be more costly than just giving the pills to everyone.

Julia Galef: Yeah.

David Roodman: What was done in this experiment was what's called mass deworming. That's the major kind of intervention that is carried out by some of the charities that GiveWell recommends that are involved in deworming. You don't take the time to figure out who's got the worms, who doesn't. You just deworm everybody, you give to all kids in school.

What they found is that in the schools where deworming was done, kids came to school more. And-

Julia Galef: How much more?

David Roodman: Oh, gosh. I think it was something like ... attendance rates went up by something like six percentage points. Don't quote me, as it were.

Julia Galef: That's fine.

I'm just trying to get the rough order of magnitude.

David Roodman: Yeah, I'm not good at remembering a specific numbers, but it's-

Julia Galef: But it wasn't like 50% or more?

David Roodman: No, it was on the order of like five or six percentage points more, where the average might have been roughly 50 or 60% attendance.

Julia Galef: Oh, I see. Starting out.

David Roodman: Right. As I said, 50 or 60 or maybe 70% of the kids who were officially enrolled in the school would show up on any given day. But we got a five or six point bump from deworming.

Julia Galef: Why is it so low?

David Roodman: Well-

Julia Galef: Is it like illness, or ...

David Roodman: It can be illness, it could be having to work on the farm. It could be enrollment records not being reliable. There could be a lot of different factors involved.

The answer is, I don't know a lot about the concrete realities there.

Julia Galef: That's fine.

David Roodman: They also looked at whether deworming improved test scores. Did these kids seem to improve in their knowledge? And the answer was "No".

The big headline finding was about school attendance. And they also found some interesting patterns that I don't find super exciting. It's not that I mean to be critical. It's just that I think they're more exciting to economists than to GiveWell -- which is that, there were externalities, what you might call spillovers. Kids who were within a kilometer of a school that got the deworming also seemed to improve, even if they themselves didn't get the deworming.

I think they avoided giving the pills to girls over 13 because they're not recommended for pregnant women. And a girl over 13, there's some probability that she's pregnant. They can then look at whether girls over 13 also saw an improvement, even though they weren't getting the pills. And the answer was "Yes".

Julia Galef: And the spillover effects, it's not just magic. It's about contagion, right?

David Roodman: That's the theory. The data don't tell you the mechanism, but it makes a lot of sense. Yes, if your friends and your siblings have been dewormed, then you're less likely to get worms, too.

Julia Galef: Nice.

David Roodman: Yes.

There were these spillovers. That was seen as very central to the paper.

Julia Galef: Nice.

David Roodman: One thing that did mean is, if you want to look at the impacts of deworming, you can't randomize within schools. You can't give deworming pills to half the kids because you may be helping your control group almost as much as the treatment group.

Julia Galef: That's a great point. Yeah.

David Roodman: If you look at the difference, it's gonna be very small and you're going to underestimate the real benefit. That shows a necessity for when you're dealing with highly infectious disease, for doing what they call cluster randomized studies, where you don't randomize it at the individual level, but you randomize it at a school level, or the regional level. Some schools get the treatment and others don't.

Julia Galef: Nice.

David Roodman: One critical question you can raise is about whether improving school attendance is really such a good thing. My former colleague Lant Pritchett is a brilliant development economist, has written a lot about how schooling ain't learning. I think maybe he's even got a book or paper with that title, saying we shouldn't be getting too excited just because you packed the kids into the schools, unless they're actually learning more -- and a lot of kids, they're not.

Julia Galef: Well, or unless completing your degree makes you more able to get a job.

David Roodman: Sure, that's right. Yeah. But I don't think we can take a lot of consolation in a child finishing second grade if he or she can't even read, you know. That's the kind of level we're talking about here, in terms of the concern.

Julia Galef: Actually, this is another brief tangent. I'll probably regret this as the end of the episode approaches and I haven't gotten to all my other central questions, but I was reading an interesting debate on Facebook recently about this basic income study. I forget in which country. It was an African country.

David Roodman: Probably Kenya.

Julia Galef: And the researcher was ... or maybe it was just a journalist. I'm sorry, I can't remember the details. But basically they went around and asked people what they were spending this unconditional cash grant on. And one of, or maybe several of the respondents said, "I'm spending it on sending my kids to school, because I couldn't afford to send them to school before." And this seems great, but, my friend said, "Actually, this isn't great because schooling is kind of a positional good." It means your children are going to be more likely to succeed relative to other people, but it doesn't actually make the whole country better off. Does that make sense to you?

David Roodman: That is certainly potentially true. At the end of the day, it's a question of context. If the schooling is effective and its graduates are entering an economy that's fairly dynamic in creating more opportunities where their improved skills will be rewarded, then I think more education will be a good thing.

But if it's really just a waste of time and they're not learning, but they're positioning themselves relative to others, then yeah. Then that benefit can be much lower than the gross benefit for the kids who are getting ahead.

Julia Galef: Okay. Cool. I'll stop derailing you.

David Roodman: No, this is all very interesting.

This study came out and it was very influential within economics in several ways, as we've discussed. But it absolutely is true that it created a lot of excitement around mass deworming, and that it has been controversial, because it is not the only study that has been done of the impacts of these deworming pills. There have been several dozen studies done over the last few decades.

Julia Galef: Wasn't there a follow up study several years later, looking at the earnings of the children who had received the deworming versus those who didn't?

David Roodman: That's right.

The same authors, plus some new guys and gals, did a follow up study where they followed a random fraction of the kids in the original study into late adolescence, even early adulthood, and wanted to know how they're doing. Was there long term benefit from getting a few years of deworming back in your youth?

And they looked at a lot of different things. They looked at income and health and education levels. There's an array of findings, but the one that popped out most, certainly for Givewell, was that former children who had received the extra deworming were earning more in adulthood. I think it was about 12% more overall if you focused on wage income, which is one type of income.

Julia Galef: 12% extra earnings seems like a lot to result from a 6%, if it is roughly 6%, increase in school attendance.

David Roodman: Maybe so. I don't really know--

Julia Galef: Although maybe that's not the only mechanism.

David Roodman: Yes. Right. As they say, I don't have priors about that, but that raises a big question, which is, "Is this plausible?"

Julia Galef: Right.

David Roodman: Right. Is it plausible that you can have an arguably modest benefit in the short run and what looks like a larger benefit in the long run?

Julia Galef: Right. Yeah.

David Roodman: And at some point you have to say, "Well, whatever the evidence said happened is the truth. We shouldn't go too far in imposing our priors."

Julia Galef: Depends on how strong the evidence is.

David Roodman: That's right. Which is always a complicated question. This is a very important finding for GiveWell, because that bump in income is huge if it lasts many years, say 10 or 20 years or more, and you compare it to the cost of achieving it, which was just a few dollars. These pills cost almost nothing and mass administration is very cheap.

You can get a benefit cost ratio easily of a 100 to 1, or a 1000 to 1, depending on how you do the numbers. That is part of what's really drawn GiveWell to deworming charities.

Julia Galef: Yeah. Plus, you have the additional non-negligible benefit of and these people don't have to suffer from worms for years. That's also nice.

David Roodman: Absolutely.

Julia Galef: Yeah. Let's get to the wars themselves. I really liked ... You had a line in one of your blog posts about your analysis of this issue that went, "And so on the dreary plains of academia did the great worm wars begin," which I felt really captures the excitement of the topic.

David Roodman: Yeah. At least for a few, maybe for you and me. We'll see if it captures it for everybody else.

Julia Galef: What kicked off the wars? Where did the doubts about the findings begin?

David Roodman: There's an institution in the UK called Cochrane. I think it was formerly called the Cochrane Collaborative. It supports a lot of the work that is called meta analysis. That’s when you're wondering whether screening at age 45 for prostate cancer actually saves lives. And maybe there've been a dozen studies of this done over the years in different ways and you want to draw them together and try to synthesize a single conclusion out of them because the larger your study, the more precise your answer can be on any question. If you can pool a bunch of studies, you're essentially making one big study with more statistical power. This is what meta analysis does.

Working under the Cochrane umbrella, people have been maintaining meta analysis of studies of the impact of deworming. I think the latest of the iteration was released ... I think it was 2015, right? Is that what I said in the blog post?

Julia Galef: I don't remember, but sure.

David Roodman: Yeah.

Looking at a couple dozen studies of mass deworming, as well as targeted deworming, just for kids who have tested positive, there was substantial evidence of no impact.

Therefore, the implication was that it was crazy to be spending hundreds of millions of dollars deworming kids when all the evidence seems to say, "Well, there is no benefit here." That has led to a kind of a classic academic argument where there's back and forth. The authors of the original studies have replied and the critics have added new arrows to their quiver. They've done replications of the study. They were just talking about and raised doubts about the methodology. It gets fairly complicated.

I think one of the big divisions here is between epidemiology and economics. To the extent it seems to work out to be between Brits and Americans. There's some of the tribal feel to it.

Julia Galef: Do the Brit and American camps line up with the epidemiology and the econ camps? Or are those two separate axies?

David Roodman: In this case, pretty much. Yeah.

Julia Galef: Interesting.

David Roodman: They pretty much line up. Most of the critics who are really saying that they shouldn't believe these deworming impacts, that is at least you shouldn't be funding deworming, are British epidemiologists. Most of the supporters of this are American economists. It's a funny thing.

Julia Galef: Interesting. Okay. Can you unpack why the battle lines got drawn that way?

David Roodman: Like your earlier historical question--

Julia Galef: I know!

David Roodman: ... I've warned you about.

I think one thing that's going on is ... I guess it's an epistemological question, "What constitutes valid evidence of impact?" It seems to me that what a lot of what the epidemiologists are saying is, "We want our statistical methods

and we get some kind of estimate of the impact, which is usually positive. But when we look at the 95% confidence interval around it, it includes zero."

I can expand on what that means. Therefore, there's no evidence of impact. They're saying that on the evidence that we have ... I should articulate this better. We think that if there were no impact, the actual data that we got don't seem terribly unlikely. When we reject to something at 0.05 what we're saying is that if there's no effect, if the true answer is zero, that there's a more than a 5% chance ...

Julia Galef: So if you're rejecting the valid 0.05 then we think there's a less than 5% chance that we would've gotten results like this? Or more extreme than this.

David Roodman: Yeah.

Julia Galef: In this world in which there was actually no effect, right?

David Roodman: Yeah.

Julia Galef: It's a surprisingly complicated concept to, even for people who are very familiar with it, to explain on the fly.

David Roodman: Yeah. As you can see, I don't actually articulate it that often or I'd have more practice.

Yes. They're saying that they cannot reject a null of no impact.

Julia Galef: Yeah.

David Roodman: They're saying that we're not confident enough that there's an impact here that we can, when forced to make a binary call, whether it's present or absent, that we can say that it's present.

Julia Galef: Yeah. And what about the American economists?

David Roodman: They say a couple things. What I might say is what GiveWell ... I think I might contrast that first with how GiveWell thinks about things.

GiveWell is trying to do a cost benefit analysis of deworming, right? And not just deworming, but deworming charities. How much does it cost to give a kid a pill? Is it $0.12? Is it $0.27? Whatever. Then, what are the expected benefits from that? If we do our best analysis, we're not gonna know for sure what the benefits are because all the studies have uncertainty in them. We can imagine at least informally that there's some kind of probability distribution. Maybe for $0.27 we'll get $100 of benefit, maybe somewhat less likely it will be $1,000 or maybe it'll be $10. There's a range or there's a probability distribution. And whether 95% of that distribution is to the right

of zero, we don't really care about. We're just trying to make our best guess and then go with it. And of course you can revise that best guess over time.

What I think is important to understand there is that we're going from questions about what is the state of the world, what is the impact of deworming, to questions about what should we do or what should visitors to our website do. So there's a jump there from conclusions about the world to decisions about what should be done.

Julia Galef: Is this sort of like Spock saying, "Well, we don't have enough data to make a decision about which path we should take." And the Captain being like, "Well, we gotta take a path, so lets just go with one of these. Whatever is your best guess." And Spock is like, "No, we don't have enough data. We can't decide."

David Roodman: It is kind of like that. That's right. You have to make a decision--

Julia Galef: It's a caricature, but ...

David Roodman: Right. Exactly. You have to make a decision based on the evidence that you have. And what your decision frame is, what your choices are, what are the downsides if you're wrong… Should influence whether you want to impose as 0.05 as your threshold or 0.2 or 0.7, or get rid of a threshold altogether and just deal with the continuum of possibilities.

What I think ... one thing that epidemiologists are doing is, because they're so steeped into the methodology of Cochrane and meta analysis, is that they're thinking inside the world of conclusions. That is, "What do we know about the world?", not thinking about the complexity that is the decision making. They feel like their job is to set up a very high bar for demonstrating that something actually has an impact.

It reminds me of an example. If you and I are driving down a highway really fast, and it's dark and rainy, and you're driving and I'm riding and I say, "You know what, Julia? I think we might be headed for a cliff. But, I can only ..." I won't worry about the language of the null. "I'm only 94% sure, so I can't reject it at 0.05." And you can say, "Well, in that case. I'll keep going."

Which would be crazy, right?

Julia Galef: Yeah, although presumably, it sounds like you and I think that GiveWell, I don't want to speak for them, did not ... if you had to guess, you would think the probability that deworming helps is more than 5%, significantly more, right?

David Roodman: That's right, yes.

Julia Galef: You're giving this sort of conservative case. That, let's say we were only 5% sure, should we still go ahead and do it if we thought that benefit was great enough? And that is an important question, but in this particular case it sounds like there was actually reason to be more optimistic than that.

David Roodman: Yes. I think we try to get away from the question whether it does or does not have an impact, because reality is not binary. What we're dealing with here, at least. Instead, think about what is our best estimate for the expected or average impact.

Julia Galef: Right. Yes.

David Roodman: And my point is this, how you synthesize the evidence into a decision has to depend on the specifics of the decision you're making. It's one thing if you're thinking about trying an experimental new drug on a cancer patient. It's another if you're worried about whether you're heading for a cliff. Right?

I could say that epidemiology is conservative and they insist on that 0.05 as a high standard of evidence to meet because there are good reasons, like maybe a lot of research is funded by drug companies and so you should start from a standpoint of strong skepticism. Or maybe most new things have side effects that you need to worry about. And again, that argues for a high standard.

But here we've got something that, apparently there's a consensus, has very little in the way of side effects and it's very cheap to deliver. And so I just think the point of high standard doesn't make sense there.

Julia Galef: Yeah.

David Roodman: Now, the opponents also argue that if you do the meta analysis right that you can meet the 0.05 standard and that's something that we can also talk about, but it's complicated.

Julia Galef: Yeah, well maybe ... feel free to go only as far into the weeds as you think is appropriate here, but if we could just briefly touch on the nature of the criticisms of the original deworming study: Maybe a good way to divide it up is ... Actually, in our last episode together, I think we talked about internal versus external validity, which is a distinction I find myself using a lot. Basically, internal validity is about, "Was the study well constructed? Can we trust the results of this study in the context where it was conducted?"

And then external validity is basically about how much can we generalize the results of this study to other contexts?

I find myself making this distinction colloquially when, for example, someone is giving me advice, and I try to separately ask, "Do I think their interpretation of their own situation is sound?" And then, "Do I think their

advice applies to my situation?" Anyway, I find that concept useful, both in evaluating studies and in life.

But getting back to the topic at hand, do you think that concerns about the deworming study in Kenya, the Miguel and Kremer study, were more about its internal validity, like it wasn't well done, or that it's results didn't apply to other situations?

David Roodman: That's a great framing.

I think a lot of the great criticism was about the first, whether the study was actually well done and is something you can believe. One can certainly also talk about external validity -- because we talked before about, "What percentage of kids have worms in general?" And I mentioned that it was particularly high where this study was done. So then you can say, "Well, that means maybe it doesn't generalize," which would be something of a valid concern.

Julia Galef: Because in other areas, if you give everyone deworming pills and only a few kids needed them, then it's not very cost effective.

David Roodman: Right.

Julia Galef: Yeah.

David Roodman: It turns out, and this is a funny thing, as I mentioned, this is the study that launched the randomization movement and development economics -- and it wasn't randomized.

Julia Galef: It was pseudo-randomized?

David Roodman: I guess, yeah, you might ... we often use pseudo or quasi as a prefix in this kind of situation.

Julia Galef: It was randomized-ish.

David Roodman: Randomized-ish. Yes.

Maybe there's certain people who use the word randomized to embrace what was done, but they did not flip a coin or roll a die or use a computer to generate random numbers. I think they wanted to, but the economists themselves didn't actually run the experiment. They had to work with another organization, an NGO, a non-profit that was operating there. Apparently, the non-profit was resistant to the idea. Some people have moral concerns about randomization. It feels like you're playing God with health and life.

Julia Galef: Although you still have to divide up the schools somehow. Whatever method you use, you're denying potentially useful treatment from half the students, whether you made that decision by flipping a coin or by drawing a line in the alphabetical roster of schools, I don't see how that matters.

David Roodman: Right. I agree, and I would also argue that we should get the best evidence that with can so that we can help [more] children in the world.

Julia Galef: I mean, that too, yes!

David Roodman: Anyway, what happened is they made a list of 75 schools that they were going to work in and then they sorted the list and it turns out that they didn't even describe correctly how they sorted it. They originally said it was by the name of the school, which was not true. It was by county, equivalent of a county. And then within counties they sorted by the number of kids in the school. Once they had this sorted list, they went down the list and they divided the schools into three groups, kind of like ... I was gonna say duck, duck, goose, but I think that's not the right analogy.

So the first school on the list went into group one. The next school went into group two. The third went into group three, and then the next one went back into group one.

Julia Galef: Okay.

David Roodman: They randomly ... not randomly, but *arbitrarily* split these 75 schools into 3 groups of 25 each. Then group one immediately got the deworming pills. This was in the beginning of 1998. Group two didn't get deworming pills until the beginning of 1999. And group three was the full control group. It didn't get the pills for the duration of the experiment, which was two or three years. I guess three years.

There was that. And you could say that this was in fact not randomized and maybe there are stories that would explain the results that they got, without relying on the idea that the pills made a difference, because of the somewhat non-random way that these schools were grouped together.

I spent some time trying to make up that kind of story. I thought, "Well, maybe just by chance the schools that got deworming earlier were at a different elevation." This is a hilly area, and that would affect the prevalence of worms, there was a big El Nino event at the time so there was a lot of flooding, which could have aggravated worm prevalence in some areas more than others. Maybe you could sort of construct a story that ... I went to great lengths to try to come up with things like this and prove them, and ultimately failed.

I really tried to attack the premise here. Bringing in new variables that haven't been checked for statistical imbalance. And mostly I came away convinced that while yes, there was this asterisk on this study, that it didn't

seem like a very compelling explanation for the results, compared to the more straightforward explanation that the pills actually made a difference.

Julia Galef: I really liked your explanation of how to approach the question of, "Are these results real?" Or "How real are these results?", in that most people who are bothering to critique methodology, which is a small subset of people, they approach that question by tallying up flaws in a study and asking themselves, "Is this study too flawed or not too flawed? Or, which of these studies is least flawed?"

By contrast, you are asking what I think is a much more important or central question, which is, "Looking at all this evidence together, which hypothesis about the world makes this entire body of evidence most likely?" Just to simplify it, is it the hypothesis that deworming does work, does reduce poverty over time, or not?

I think you described this process as kind of related to Occam's razor, right?

David Roodman: That's right. When you're trying to make decisions whether to recommend a charity or not, it can't be optimal to set such high standards for the quality of the research that you have no evidence and you just say, "Well, we don't know." Because that's like Captain Kirk and Spock, as you described.

We have to just say, "Here's what we've got. We gotta make the best of it. What's the best explanation for all the data?" And I define best as being a combination of two things.

A good theory is something that explains most of the evidence before us and is simple. That's what is called Occam's razor. That's the idea that the simpler theory ... if you have two theories that can explain the results before you, the simpler one is more likely to be right. That may reflect my biases. What I'm actually trained in is not economics but mathematics. In mathematics, the simpler statements are more beautiful and more true in some deep sense.

You raise an important question, which is, "How good is a study have to be for us to take it seriously?" One of the things that epidemiologists say about a lot of the research that economists have undertaken is that it doesn't meet good quality standards. For example, in Kenya, the children in the control group did not receive placebo pills. They knew they were in the control group, whereas the kids who got the pills knew that they were in the treatment group. That is not considered best practice in medical research, because then you've got some people who are in the treatment group who know they're being observed and are being treated differently and that can create, what is it called, a Hawthorne effects and placebo effects. By virtue of people knowing they're being observed, they can behave differently.

Julia Galef: Right.

David Roodman: And the analysis was not pre-registered. Once you've got your data in, you can test all sort of hypotheses and then maybe there's a tendency to only report the results that are significant.

Julia Galef: Right.

David Roodman: So you're data mining. One way to prevent that is to pre register what you're going to do with the data once you get it. And that was not done here. And so on. There are a bunch of these things.

Whereas in epidemiology, I just say in medical research it is more common for there to be pre registration and blinding and double blinding and these kinds of things.

I intuitively reacted against this kind of binarization of knowledge, saying it either counts or doesn't count, it's either there or not.

Julia Galef: There's two different thresholds here. One is the, "Is it significant, or is it not significant?" There's just a binary cut-off. And then another is, "Is the methodology good enough, or is it not good enough?"

David Roodman: Exactly right.

But I realized that I also do the same thing. It's like accusing somebody else of being racist and then realizing that you're also superficial in how you judge people. There are some studies that I would hardly spend any time looking at because I just don't trust them. If you tell me that you've done a non-randomized study where you discovered that healthier kids get higher grades, I'm not gonna take that as proof of that either causes the other.

Julia Galef: But that seems correct to me.

David Roodman: If the only evidence we had on something was of that quality then maybe we should just not pay attention to it.

Anyways, I filter too and the question is, when is that appropriate and how should you do it?

There was a sort of a holier than thou kind of thing coming from the epidemiologists. And the economists responded by ... everybody agrees that blinding would be good, but the economists have good responses to why at least in this case they couldn't do everything that the epidemiologists said would be ideal. To get a bit graphic, if you have worms and you take the pills, then your body will expel worms and you'll see it.

Julia Galef: Yeah.

David Roodman: So you can't actually hide from the subjects who's in the treatment group and who's in the control group. It's just not possible, that kind of thing. So blinding was not actually possible in this case.

I forget how we got into this, but this is an important thing that I talk about in the question of whether the study itself is valid. I agree that in the ideal, there are ways in which it could have been done better, but we still have to work with what we've got.

Julia Galef: Yeah. I think that's a message that's important that often gets lost in the, overall very good and useful, message that skeptics and science communicators and statistics promoters like me promote. Which is that these are all really important elements of rigor, and you should be less trusting of a study if it didn't use blinding or randomization, et cetera. But the takeaway that people often get from that is, "This is my excuse to ignore anything that doesn't meet every single one of these standards."

Honestly, even for the gold standard, randomized, controlled, blinded, long term study in the field, there's always problems that you're gonna be able to find. And especially, I mean I'm not claiming the epidemiologists were motivated in this case to reject this result, but if they had been, it would have been easy even for the best conducted study, I'm sure they would have been able to find reasons to reject it.

David Roodman: I think that's right, yes.

Julia Galef: David, when you zoom out and you look at the way that this debate went down, how does it make you feel about the field's, or this cluster of fields', social epistemology? Do you think that people basically handled it well, that they discussed these various epistemological methodological questions in an open and intellectually honest way? Or did things become tribal with people arguing for their side and just digging in their heels?

David Roodman: I think I am a bit biased. I have much more affinity with the American economists, even though technically I'm not one. It feels to me like the economists have been more careful to be constructive in their tone and more sophisticated in how they reason. For example, not quite getting so caught in the trap of insisting on significance of 0.05 and only considering studies that meet certain superficial quality criteria.

I feel when I think of there being tribal elements to this, I feel that more in the British epidemiologists that I've talked with. But that may just reflect my bias.

Julia Galef: I'm sure the epidemiologists say, "I'm not tribal. You're tribal."

David Roodman: Yes, and they might have some validity in saying that.

I think that the meta analysis process, which is supported by Cochrane and also another group in the United States called Campbell, which has been an important part of this whole debate, is still fundamentally constructive. I think there are conservative biases, as we talked about towards really imposing high standards for saying that there's an impact, which may not be appropriate here, but I still think that their general approach has been useful.

Where I feel like the biggest problem lies is in people being systematic in their reasoning -- and this may be very familiar territory for you. There are times when I was debating with some of the British epidemiologists in email where I felt like we weren't quite talking past each other, but we were operating from different premises and we were not doing a good job of confronting those differences and working with them systematically.

Julia Galef: This is why I like to have a whiteboard when I'm debating something even remotely complex, so that we can map out: What are the things that I'm assuming that makes me say this? And can we pinpoint where our models start to diverge?

David Roodman: Right.

I think what the epidemiologists are saying is that generally when we run the meta analysis we do not get the results that are [significant] for the impact on children's weight, for example. They are not statistically significant at 0.05. Therefore, there's no impact. Therefore, people who are engaging in deworming on the basis of the evidence are highly misguided and they're wasting money. Maybe even doing harm.

There's a set of assumptions there and sort of established ways of reasoning that I don't think are universally shared, at least not with economists.

This is a case where ... we're not talking about “what is art,” where at some point you can argue and never reach a conclusion. You're talking here about data, probabilistic reasoning, making concrete decisions. And there actually are ways to be more systematic about that.

And so I've been playing now with building a spreadsheet where you can engage in some more Bayesian and systematic decision making. You could say, "Here's my prior for what I think the impact of mass deworming is on children's weight, which is an indicator of nutrition. Here's what the evidence says from the meta analysis, and then I'll do the Bayesian analysis, which produces a new distribution for the estimated impacts that factors in the data but also my preconceptions, which are unavoidable."

And then I can feed that into some kind of cost benefit calculation, where I say, "Here's the distribution of likely benefits based on the evidence we have and here's our best estimate of the cost of the pills, here's a discount for the

fact that where deworming is done today, if your kids have worms, expect those benefits. Here's other discounts, et cetera."

Julia Galef: Wow, that's better than my whiteboard solution.

David Roodman: Ideally, they'd be fully graphical and you can adjust some knobs, and you could play with your assumptions. Maybe a key assumption, for example, would be that there's no side effects. Maybe if you start to allow side effects, that'll really shift your decision.

And then at the end of that, you would get a distribution for the cost benefit ration.

If GiveWell is right, the center of that distribution will be positive.

Julia Galef: Right. And you can just decide, "Is enough of this distribution positive enough that it's worth taking action?"

David Roodman: Right. And I think as a matter of math and formal analysis, that encompasses both sides here.

I'm making this up as I go. I've not articulated this before. But I ought to be able to say to the epidemiologists, and the economists, "Do you agree with this structure here in the parameters? If you don't like the parameters, the spreadsheet lets you change them. If you don't like the structure, tell me, and we can redo this."

Julia Galef: Yes.

David Roodman: But I feel like it could provide a common ground for more systematic discussion of how we go from the evidence that we have to the decisions that we need to make.

Julia Galef: Yeah. I mean, that's basically my ideal ... it's funny. I think people are used to thinking of, "Well, lets just check the data and see what it says" as the end point. That's how we resolve the debate, is we just look at the evidence. But in many cases that's just the starting point.

David Roodman: That's right. That's because in many cases the evidence is weak, there's a lot of uncertainty, or it comes from one context and we're working in another and so-

Julia Galef: And you have different priors, and you have different epistemological priors, about how you should interpret data.

David Roodman: Yeah. If the evidence is really powerful then most of this falls away. We have enough evidence that smoking is bad for you, or that if you run into a wall, it

will hurt. But when you move into more uncertain areas, then you have to be more careful in how you go from inference to decision.

Julia Galef: That seems like a good place to wrap up the conversation. Before we close the episode, David, I want to give you an opportunity to introduce the Rationally Speaking pick of the episode. This is book or article or something that has influenced your thinking in some way. Do you have a pick for us?

David Roodman: Yeah. When I first got interested in public policy and economics in development and other things, I was just out of college, trying to figure out what I wanted to do with myself. I was actually studying in Cambridge in England and I lost interest in mathematics that I had been studying there, and ultimately actually failed all my exams. I was-

Julia Galef: Oh, no.

David Roodman: I mean, I just stopped attending classes. But part of what pulled me away from lectures was I was starting to read books that friends recommended to me or that I found on their shelves. One of the most exciting books for me in that period was called Steady-State Economics by Herman Daly, which brought notions of economics but also thermodynamics to thinking about global environmental issues in a way that I found extremely compelling and helped me put me on the course that my life eventually followed.

Julia Galef: That's great. One of the categories of books that I love is a book that highlights a parallel or similarity between two fields that you thought were unrelated and says, "We can use these tools of analysis or these ways of thinking from one field like thermodynamics and this other completely seemingly unrelated field, like economics."

David Roodman: Yes.

Julia Galef: Cool. Well, David, thank you so much for coming back on the show. It's been a pleasure having you.

David Roodman: It's been a pleasure to be here. I look forward to doing it again.

Julia Galef: Yes. Same.

This concludes another episode of Rationally Speaking. Join us next time for more explorations on the border lands between reason and nonsense.