

Rationally Speaking #217: Aviv Ovadya on “The problem of false, biased, and artificial news”

Julia: Welcome to Rationally Speaking, the podcast where we explore the borderlands between reason and nonsense. I'm your host, Julia Galef, and I'm here with today's guest, Aviv Ovadya.

Aviv is the former Chief Technologist for the Center for Social Media Responsibility at the University of Michigan, and for the last couple of years now he's been working on tackling the problem of misinformation, disinformation, and other related threats to our democracy and the stability of our civilization.

So that's what we're going to be talking about today. Aviv, welcome to Rationally Speaking.

Aviv: Thanks for having me on.

Julia: So, Aviv, there's this very widely shared article about you in BuzzFeed earlier this year, and it described how you had sort of tried to sound the alarm about what they referred to as fake news to Silicon Valley, pretty early on, before the 2016 election. And your warnings were, as they said, not taken all that seriously.

Can you describe, in what way did you sound the alarm, and how did people react?

Aviv: I talked publicly and privately to people at the different platforms around how the information ecosystem was being weaponized. And in particular, how the incentives resulting from the way the platforms were working at the time was leading to sort of a rapid increase in the amount of misinformation that was being shared online.

Julia: And by "the incentives" — you mean misinformation was likely to be shared because it was sensational, or people got really worked up about it or something? And so since that made it more likely to be shared, that created an incentive to produce misinformation?

Aviv: Yeah. So that's one huge component of it, just the fact that if something is sensational, you're more likely to share it. But there's also the recommendation engines themselves reward that sort of behavior.

Julia: Right. And in what sense were they not all that compelled by the warning?

Aviv: Well, there's always this, "We aren't responsible for our impacts," or "If we're do anything it'll look like we're biased." Or, "This is really hard." All of the sort of standard answers. Or like, "This is just not what we do."

Julia: Right. So it was less about them denying that this pattern existed, and more about, "This is not our department, or not tractable for us to tackle."

Aviv: Where there also wasn't awareness, again, of the sort of scale.

Julia: I see.

Aviv: And so, anyone can say, "Oh, this isn't something that we need to focus on, this isn't our problem," when something is small. But once something is ginormous-

Julia: Go for it, yes — that is or should be a word.

Aviv: ... yeah, it's sort of a different story.

Julia: Yeah. So this term "fake news" that the BuzzFeed article about you used, and I guess is in common parlance now, seems to me that it encompasses a bunch of different things. So to talk separately about the different things under the fake news term it seems to me there's three categories:

Where category one is false articles, articles that are saying something clearly, unambiguously false. There was an article going around in 2016 that the Pope had endorsed Donald Trump, that was definitely false, verifiably false.

Aviv: Yes, many, on the order of a million shares.

Julia: Oh wow.

Aviv: Or engagements at least.

Julia: And I would also, in this false news category, count things that are technically labeled as satire but are clearly intended to fool people, especially when shared out of context without the satire tag clearly visible.

So that's category one. And then category two would be biased or misleading news, which is arguably a majority of news. I don't know, Fox News disproportionately covering or more harshly covering crimes by immigrants, or the liberal media talking about something really bad that Donald Trump did and making it sound like it's a new thing, unique to Trump, but maybe all presidents have done it too.

And then category three would be what I would call artificial news, so deep fakes. Videos that make it look like Obama's announcing that he's gonna confiscate everyone's guns or something. I made that example up, but ... and this I guess is not super common yet, but on the horizon.

So I'm curious if you are concerned about one or two of those categories specifically, or all of them. And feel free to dispute my categorization.

Aviv: Yeah, I think that's a decent way of sort of breaking down the problem area. I mean there's also, specifically this weaponized disinformation used by foreign actors, which you can throw into that mix. In addition to the stuff that's weaponized internally.

Julia: Right, it's like another axis.

Aviv: Right, and then there's just sort of rumors that kind of spread. And so that, for example, we're seeing people being killed in villages in India because there are

rumors that there's all these child kidnappers around, and so some people come in and give food to children, and then they're just mobbed on by thousands of people, and killed. So those, that type of ... in different environments there's different ways in which misinformation/disinformation can manifest.

Julia: Can you give an example of what your concerns are for each category? Or pick the category you're most concerned about?

Aviv: I think they all blend together in many ways. They all impact our ability to make an accurate model of the world, both as individuals and as a society. Or as governance, as decision makers, more broadly. Whether it's health misinformation, meaning that you don't vaccinate your kids, or that you take some weird supplements that really are not good for you. Or if it's political misinformation, that leads you to elect leaders who are just a) not going to be competent, and b) get you into major wars. There's these very, very broad ways in which misinformation can negatively impact people on both a personal and a societal level.

Julia: Does one of the categories - false news, biased news, or artificial news - seem more tractable? Or at all tractable?

Aviv: I mean false news is the most tractable in some ways, because you can just say, "Oh, this is false, here is why, here's some evidence for it." And if you come from the "There is no truth," then okay, it becomes a little bit harder. But that isn't really a practical way to approach life or civilization building. And so most decision makers don't do that, and neither do most humans, in terms of their day to day life.

So false news is in a sense the most tractable.

Julia: I guess there's identifying it, and then there's spreading the word that it's false in a way that is compelling to people.

Aviv: Right, and this all really depends on sort of the communication infrastructure that we have. So if you're working in an environment like Facebook where you are centrally controlled, and we have all the data on everything you do all the time, that's let's say an easier way to ... it's much easier to sort of handle false information, especially if it's something that is actually so dangerous that it threatens societal stability or it could lead to a World War III situation. Versus if you have some sort of super de-centralized network of information sharing where everything is encrypted, and in that environment will actually make it very very difficult to actually address any sort of false news, even if it is societally or existentially de-stabilizing.

Julia: Yeah, so I definitely, at some point in this conversation, want to touch on the question of, how easy is it, in fact, how tractable is it, to decide what actually is false news? Or to separate normal biased news from overly, unfairly biased news, that kind of thing.

But it sounded like you were about to talk about synthetic, or what I was calling artificial news. I'm curious first to hear how tractable that seems to you.

Aviv: So the synthetic media, just to be clear about what that means: so AI advances allow us to get closer and closer to mimicking reality without actually having real footage of something. So I can, for example, make anyone appear to be saying anything. Or make it appear as if any particular person is taking any action.

So you can let's say film your friend doing something, and then replace it, or put Obama in that. And we don't have the technology to do all this super well yet, but we're very very rapidly developing that capability. And there are particular ways in which we can already do it, sort of.

So that, the decrease in the barrier to entry to doing that sort of video manipulation - 'cause without requiring a production budget of a million dollars, to pull off something - just enables a broader set of actors to both weaponize this and also to just emergently end up being confused and tricked. Or enables not only people to fool others at scale, and in a way that's much more compelling than just written media.

But it also enables people to say that thing that actually is real is not real.

Julia: Right.

Aviv: And I think that's even, in many ways, the deeper threat, is that you lose your ability to hold people accountable in ways that are potentially necessary for certain levels of societal function.

Julia: Like Trump saying that the ... what was the name of that video where he talked about grabbing women?

Aviv: Oh, Access Hollywood?

Julia: Access Hollywood, yeah. Saying that the voice had been manipulated or something like that. That is a ... it was not true in this case, but it's a plausible claim in this age.

Aviv: Yeah, and thankfully the technology is not currently out to be able to do that sort of synthetic creation, at least out of the lab. But being able to use such evidence is crucial, both in the court of public opinion and then in the courts more broadly.

Julia: So I was kind of assuming/hoping all this time that verification technology would just keep pace with falsification technology. And that we could ... there could just be a little browser plug-in, or just a feature baked into Facebook or Twitter or whatever, that would have a little red "false" tag, or "fake," "synthetic" tag, in the corner of videos that are detectable as being deepfakes. Are you not as optimistic as I am about that?

Aviv: Well that requires a lot of things falling into place. First of all, this is a cat and mouse game, right? So you're gonna have your fakers who are, if they're able to also use your detectors, or if there is some kind of real time detection, then they can train on your detector.

Julia: Right.

Aviv: So based off the speed of your detector they can just be like, "Oh, let's train our system so that it specifically fools that detector."

Julia: Ah.

Aviv: And so, there is some sort of disagreement within the research community around whether or not this is tractable in general, whether or not the sort of attackers always win, whether or not defenders always win. My stance is that the attackers will win a significant proportion of the time in sort of the steady state. They'll always be able to be a little bit ahead of the game. And so, that doesn't lead me to be super optimistic, but it does mean that there's other infrastructures we need to be building in order to help defend our information contents.

Julia: Do you have any ideas about what that infrastructure could look like?

Aviv: So part of that is, I think of this as sort of a knowledge pipeline. Where you've got your media, which is — there's the creation of that media, there's the distribution of that media, there's the way it affects belief formation, and then the way that it affects action. At each of those trenches you need to sort of make that component more resilient to this sort of attack.

So at creation time you can have strong incentives to create timestamps and even place stamps that are auditable. And you might be able to throw in a little bit of block chain for that, but it really is not a be all and end all solution for this.

Julia: I was going to ask about blockchain, because I know I'm supposed to.

Aviv: Yeah, no. There are little sprinklings of it that you can throw in here or there to affect, or address very minor aspects.

If you, for some reason, don't trust three different independent bodies that could all be sending timestamps to you, you could put it on the blockchain and then that is even more verifiable. So yeah, for time stamping it's sort of nice.

There's maybe a few other minor areas where it could be fit in, but it's ... the core here is around how you even do that sort of time/place verification in the first place? Like, how do you deal with the analog hole, right?

Julia: The analog hole?

Aviv: The analog hole is where you're just taking a video of a screen, in a sense. So you can say something happened at a time and a place, but actually it's a created video that you just took a video of. You are taking a video of something that was on a screen anyway, so the thing didn't actually happen at that time and place.

Julia: And you're saying this makes ...

Aviv: This makes it harder to know, "Oh, this thing that you said happened at a time and place actually did." Because someone could've just pre-recorded it from any other time and place.

Julia: I see. Okay.

Aviv: And there's still, you can sort of have a lower bound, or upper bound, or whatever, but let's just say that this is part of it, that creation step is important.

But there's also the distribution step, where if you have significant disincentives to create and spread disinformation, or to create and spread, let's say, weaponized synthetic media. Then that can again increase that sort of barrier to ... or the likelihood that somebody is going to be trying to promulgate that content.

And I think you... can at a high level think about this as, you want to increase the cost of faking, and you want to decrease the cost of proving something that's actually real is real. And that's sort of the overall framework.

There's other things that might be valuable but are definitely more controversial. So for example if it's so easy that you can just take an app on your phone and record your voice as anyone else's voice, and then use that to, for example, call people and ask them for money, as their kids.

Julia: Yikes.

Aviv: Which already happens, it just doesn't happen as effectively.

Then there's some incentive to say, well maybe the App Store shouldn't let you install an app that lets you record anyone's voice without that person's permission. So only apps that sort of have gone through some sort of quality check of hey, this cannot be easily weaponized, in terms of the way it can create synthetic media.

And so, again, this is based of the fact that there is this sort of external control of what can show up in App Stores. And so it's almost the authoritarianism of the App Store is the thing you're using to prevent the sort of mal-democratization, this democratization of, "Oh, you can do all this amazing, creative work," but it may be the net impact of that is actually negative. Maybe it's even very, very significantly negative. And so you can think about this not just as democratization; anyone can do anything, but mal democratization; anyone can do anything in a way that's significantly harmful.

And then again, there's thresholds for that. And there's definitely an argument to be had about how bad that is. And I would argue that its probably, there's a likelihood that it's pretty bad.

Julia: So I know I've read that fact checking in general isn't all that effective. Correct me if I'm wrong, but I think that Facebook earlier this year or last year tried to label links that people shared as being false, or potentially false. And that researchers found this didn't have that much of an effect on whether people, I guess, shared the story? I don't know if they checked whether people believed the stories. But the takeaway in my mind from what I read was that it didn't have that much of an effect.

But do you think that we have reason to be more optimistic about fact checking of just, like, is this video synthetic or not? That seemed like the kind of thing people would be more receptive to, that they would consider more trustworthy than fact-

checking along the lines of, "Are the claims in this article roughly accurate according to me, the fact checker?"

Aviv: Yeah. I'm hopeful that there is a little bit of a difference there that makes it a little bit more receptive, at least in some domains.

Julia: Yeah.

Aviv: Because, as you say, it is slightly different to say, "This was literally not a thing that happened." Where, like, other people would be, say, that, "Oh, this is literally a thing that happened."

And you can sort of say, "Well, you can prove that that person wasn't at that place." "Okay, then this, this is not legit."

That said, motivated reasoning still does kick in, and just not trusting the people who are doing the verification does kick in.

Julia: Yeah.

Aviv: And so it's still not any sort of end all and be all.

Julia: Okay, so turning now to the other two categories of false news and biased news — you and I have had some conversations before about this, where I was a lot more pessimistic than you about how tractable it is to label and disincentivize false and biased news.

I previously called it an "AI-complete plus" problem, which our listeners ... If you call something AI-complete, it means that that problem is ... In order to solve it, you'd essentially have to create like human-level general intelligence. So people will talk about machine translation, being able to do an effective translation of a work of fiction, for example, or any kind of nuance work — preserving the connotations and the mood and all the background context. That would be an AI-complete problem.

And so I saw this issue of identifying and disincentivizing biased and false news as being AI-complete plus, in that having human level intelligence would be necessary, but not sufficient to solve it. The reason that I thought that is humans can't even agree on what — except for extreme examples, like the Pope did not endorse Donald Trump, most of us can agree on that, although, as you say, there will always be a small cadre of people who just won't buy anything you say if you're a liberal, or a conservative, or whatever.

But except for those extreme examples, it just seemed to me that no one can agree ... It's extremely rare for someone who's liberal to say, "Yes. That article is biased in a liberal direction," or someone who's conservative to admit, like, "Yes. That pro-conservative article is biased."

And so the question of tagging things as being like extremely biased or so biased they could be considered false, again, except for extreme examples, would just be ... it would just collapse into people agreeing with liberal articles if they're liberal, and conservative articles if they're conservative.

Anyway, that's the summary of my disagreement. I was just thinking about it more recently, since our last conversation and I think one crux for me, and I'm curious about your take on this, is that it seems to me that those extreme cases, the "literally false, this did not happen" cases, are not all that impactful in the grand scheme of things. That, yes, people do share articles about the Pope and Donald Trump, and so on, but they don't have that much impact on our political beliefs, or on how we vote, or things like that.

Whereas the biased news really does. Like ... I don't know ... Fox News giving people an overinflated sense of how bad immigrants are, or whether their guns are going to be confiscated. And that stuff is the stuff that's much harder to combat. So, yeah, the impactful stuff is hard to combat, and the non-impactful stuff is slightly less hard to combat.

That's my picture. Does that seem wrong to you?

Aviv: I think it is, at a high level, pretty accurate. I think that there's sort of two things I'd throw out there.

One of them is just because something is hard, doesn't mean we shouldn't attempt to do it, just because the trade-off of not doing it is so high. The other is that, even taking on those extreme cases, actually has an impact on the biased ones.

This isn't a reason to not take on the biased ones too — but if you think about the incentives here, it's a competition for attention, and it's even less ... People say, "Oh, Facebook and YouTube, whatever, are competing for attention." That's sort of true. But if you talk about who is really doing that, it's the publisher. They're doing that in the framework of these platforms.

And you can actually see this explicitly. You have the person who's just sort of taking over HBO says, We want to get as much time, as much engagement, as possible. Very explicitly talking about this in an off-the-record session that was made public. "We want to get those dopamine hits." They're explicitly trying to pull people off other platforms with the things that will drive them, and that in many cases is the ... Or in order to do that, you need to have it be more emotionally salient than even, let's say, that extreme thing.

By taking the ways in which online misinformation has sort of made that race to the bottom ... Like giving it even a new bottom, then a new bottom below that bottom, has, I think, effected the entire information ecosystem.

Julia: Can we talk about some ways that people have started trying to tackle this problem? Like, I mentioned earlier, Facebook trying to flag links as suspicious or potentially false. And earlier this year, Facebook announced that they were changing their News Feed algorithm to prioritize personal news, personal posts more than political or sort of professionally produced content. Do you know of other approaches?

And I'm also curious if you know whether Facebook's News Feed had any impact on misinformation.

Aviv: Yes, so just to clarify what's happened with Facebook... There's many, many changes, more than we could ... that would be hours of conversation to go into all of them.

Julia: Okay.

Aviv: But the high level is that explicit labels weren't super effective. But they found that just showing related content that provided additional context, and in some cases were debunks. In addition to still warning people if you're going to share something and it was previously debunked, they warn you. And if you see something in your feed that was previously debunked, it'll have additional context about why that might have been debunked, and so that's still happening. It's just not like, "Oh, this is likely to be false," being so front and center.

Julia: And is that more effective than the original?

Aviv: I believe that was what they found. And that was also what third-party researchers found.

Julia: Was it how likely people were to share it, or to click it, or-

Aviv: I don't remember the specifics of that.

Julia: Okay.

Aviv: And some of that is not public, and even if it wasn't, I don't want to mix that up.

Julia: Going back to what I was saying might be a crux of disagreement between us about the tractability of this problem... One part of what I said was that it seemed like that impact of literally false news wasn't that great. I'm sure you've read more about this than I have. What do we know so far about measuring the actual impact of false news?

Aviv: It really depends on where. If you're talking about a place like India, you can point to a number of people who are no longer alive, right, because real life mobs spurred up my misinformation killed them. I mean, that's just lynchings. That's not even counting the broad polarization doesn't have the same level of drastic impact. If you're talking about the Ukraine, you can talk about sort of the constant information war that's being waged that impacts the stability of the region, likely the physical war.

Julia: Sorry, what information war?

Aviv: Between sort of Ukraine and Russia. Ukraine is sort of where a lot of Russian tactics were originally tested out before they were applied to America ... or continuously being utilized because Russia sort of wants more and more influence in the region. They basically tried to take it over in various ways.

Julia: And what kind of misinformation are people sharing?

Aviv: Everything. I mean, I don't even know how to scope it. If you think about what happened in 2016 with the US, that's like a tiny, tiny, tiny fraction of what

Russia's interest in the Ukraine is and the type of information war that's been waged there for a number of years. If you're talking about a place like Myanmar, misinformation helped tip the scales of public opinion enough to support what's essentially ethnic cleansing, right? And it's hard to measure that type of conflict too.

Julia: I guess we can tell very plausible stories, but how do we actually measure how much it was the misinformation itself?

Aviv: Right, and that is very, very difficult. Usually the sort of work will happen many years after the fact with like a bunch of historians and, maybe in this case, data scientists working together to sort of try to piece together ... Again, it's still gonna be a story because causality is not easy to determine here. I can tell that story, and I can point to lots of scenarios around the world where there is a compelling story and where people who are living that story experience it, and where there is evidence to support different correlations. But you cannot really prove causality in the same way that you can say, "Oh, this particular thing started World War I."

I mean, just continuing to go to the US. There, we can definitely see that there is a significant amount of misinformation that's being shared, but, again, causality is hard to determine. And without some sort of rating system to actually decide what is misinformation and what isn't, you still can't really tell even that correlation very effectively because you don't know the amount that's being shared. You don't know who that's influencing if you can't say what it is.

Julia: But, I mean, researchers have done this. They've picked examples of articles that are unambiguously false, and then they've surveyed people like how many people have seen it, how many people believed it. I think there was a study that looked at ... was it likelihood to vote or ... ? It was some sort of more impactful outcome metric.

Aviv: Yeah, so there are studies that attempt to sort of wean or separate these different components. One of the challenges of them is that ... they run against this sort of, again, the credibility rating issue of like, "Well, we can look at things that are just explicitly false," but those are only a very small part of the problem.

Julia: Right.

Aviv: And no one's really done a great analysis of sort of that broad space of sort of questionable sources or things that have more bias. It's very, very hard to tell the influence of that if you can't even describe what it is or measure it, and so this is sort of a prerequisite for that sort of research.

Julia: Are there other approaches ... It seems to me, I'll say, that there are more approaches to combating the effect of misleading, false, and biased news altogether that don't rely on the evaluative step on the step of like assigning a score or identifying which stories are false. For example, approaches that make people, in general, less receptive to misleading or bias news, like approaches to reducing political partisanship.

Not that any of them are easy. I'm just saying this isn't ... They're all hard, but this one hard avenue might not be our only avenue to approaching the problem.

Aviv: Yeah, and you can make some progress on that. You can say that, "In this constrained environment, with this set of known stories that are false and true, whatever, and with this sort of priming, they were less likely to sort of believe that thing." Those are valuable, right? But if you actually want to measure the true impact of something like that, you're gonna need to see how people interact in the real world and then, again, you're back to that same problem. I wouldn't say that credibility score or ratings are a solution. They're just core infrastructure because they allow you to do that sort of measurement and study.

If you're gonna try, let's say, changes in your design of your platform that might decrease the amount that sensationalism is actually driving sharing... How do you know if it's succeeding? Because if you can't tell that scale, what stories are being shared or not... If you wanna look at the effect of a media literacy campaign, how do you tell?

If you wanna be able to immunize people, let's say, by showing them how they have been fooled... you still need to show them how they've been fooled. You need to have this list of low scored things that they've already looked at, in order to point them to that.

Julia: I see. You're saying the evaluative step that I was trying to route around in my-

Aviv: It's core infrastructure.

Julia: It's core infrastructure for testing whether or not your route-arounds are effective.

Aviv: Yes. No matter what those routes are. I mean, you can avoid doing it, but you're just not gonna have anything as meaningful to talk about. You won't really know the impact of that.

Julia: I mean, you could look at the impact of, like, measured political partisanship, in different ways, like, how negatively do the people from different political parties feel about each other? General political literacy, I guess. Probably some measures of motivated reasoning about politics.

Aviv: Yeah, no. I think there are definitely other things that are useful to measure, and I don't wanna sort of say that those are irrelevant. But if you are attempting to see how an intervention ... Well, all of those are sort of indirect and there's many, many different factors that fall into them. If you wanna be able to say, "Okay, this actually impacts the degree to which people share this type of information," then you need to actually know what this type of information is.

Julia: What do you think about the, I'll call it the hide-the-vegetables approach. So this is sort of what Upworthy was trying to do. They were like, "Well, all of this negative polarizing news is like, it has an edge in being shared because it's so outraging and sensational. And so what if we create good content that's positive and about important things, but we just make it sensational enough that people will just share it anyway?"

So, Upworthy hasn't been doing great recently, in part, thanks to Facebook's algorithm changes, but is there a version of that you think could be workable?

Aviv: I think it is valuable. I mean, there's pros and cons. I guess there's this question of, do you want to have those who are doing good work always having to sort of compete with the stuff at the bottom, or do you just want to make that bottom have just less play in general? Or in a sense, disincentivize things from being at that bottom layer?

Julia: Yeah, I guess it just comes down to how tractable each of us thinks the different hard purchases are. So, if you think it's really hard to effectively disincentivize misleading news, then Upworthy-style approaches might seem, relatively speaking, more promising.

Aviv: I mean, with the caveat that it's so much cheaper, so much easier right now to create sort of a fake Upworthy than a real Upworthy, right? And so if you think about sort of the resource allocation and how this all would play out, the real Upworthy is going to lose dramatically relative to the fake one. Unless you, again, change those incentives.

Julia: Aviv, what do you think the stakes are here? On one perspective, the stakes are, well, if we can't solve this problem, the world gets somewhat worse. There are more instances like the ones you talked about in India have of people getting lynched for things they didn't do. Politics in the U.S. gets more divisive and intractable and the parties are just extremely polarized and we don't get very much done. And that's bad, but it's not the end of the world. Is that ... You're smiling. Is that your sense of the stakes, or no?

Aviv: I think that it is closer to, I don't know about the “end of the world” explicitly, but dramatically increasing the probability that we have the end of the world.

Julia: Oh, in what sense?

Aviv: I know many of your listeners may be caring about ex-risks or catastrophic risks-

Julia: Ex-risks are existential risks, like, risks to the survival of humanity.

Aviv: Yeah. I would probably put this more in sort of the catastrophic risk camp or at least the indirect catastrophic or existential risk. Not only that, I'd put it in sort of the urgent category of that, where you have a very, very limited window.

A way to think about this is that new communications media often destabilizes societies more broadly. So, think about the printing press and the ways in which Europe was basically engulfed in war for hundreds of years. You could think about radio and the connection between that and things like World War II and the ways in which radio was able to create this sort of new type of nationalism to some extent. And then there's lots of caveats to this nuance and you know, I can't express it in the two minutes here, but new communication media affect how people organize, effect how people make sense of the world, and they affect the resulting societal stability.

We're also living in a world now where, let's say stability isn't quite as quite where it was, where individuals can have far more influence on sort of the overall stability of the world and where you have a whole bunch of really tricky challenges up ahead within the next five to 20 years that could easily derail even a

very, very well-functioning civilization. You're in this environment, and now you're making everyone dumber. You're making them less capable of handling it, both at an individual level and at a societal level.

You can think about this as, you've got your civilization driving its car down the road. And it's now starting to take LSD, and it's like seeing these hallucinations all over the place. And it's still trying to drive. There's going to be some level, some amount of LSD or some amount of like, of hallucination that you can still sort of drive without crashing. But there's going to be some level where you can't. We're just increasing that.

And as we speed that car up, as the road gets windier, as more obstacles show up on that road ... And so that's sort of the broad framework for why our ability to make sense of the world is so crucial.

Julia: That's a very vivid metaphor. I can feel my heart rate increasing.

Aviv: Well, one additional sort of even level on top of that is or relating to that, is this international cooperation is going to be really crucial to handle some of the other sort of catastrophic risks that we'd be expecting over the next five to 15 years. How does international cooperation work in a world of increasing misinformation, sensationalism, tribalism, where those who purvey misinformation are more likely to succeed or are more likely to be attacking things that aren't there than attacking things that are and actually handling real challenges? What's the chance that we can actually get a lot of the most powerful countries in the world to agree to handle a real problem if they're all just saying everyone else is bad?

And that's one of the things that has resulted from the rise of misinformation more broadly.

Julia: Is there anything that you would recommend our listeners do if they want to get involved or find out more about the problem and and how you're thinking about tackling it?

Aviv: Yeah, I definitely hope that people take this problem seriously. If you're in one of these platforms, think about how you might build your system to mitigate that, mitigate some of these challenges. If you're interested in contributing to efforts to address misinformation writ large, both in its current form and in its future synthetic media form, definitely reach out. There's many ways to contribute, to sort of addressing these challenges. Yeah. Feel free to contact me. There's a number of other organizations that are forming in this space to take these challenges on.

Julia: Do you want to give an email address or a website?

Aviv: Yeah, you can reach me at aviv.me or on Twitter at [@metaviv](https://twitter.com/metaviv).

Julia: Great. And Aviv, before I let you go, as you know, I like to ask all my guests to nominate a source, a book or a blog or other person who they have substantial disagreements with, but nevertheless, think is valuable to read or engage with. What would your pick be?

Aviv: So, I'm going to say that both Tim Wu and Danah Boyd are both amazing people, all of who I also have disagreements with specific aspects of the way in which they think about this problem. So, Tim Wu believes that breaking up the tech companies is sort of a solution in many ways while also being sort of the expert on the ways in which tech platforms work.

Danah Boyd runs Data and Society, which is an incredible organization focused on understanding how data interacts with our society, as one might imagine. Her work on the failures and media literacy is crucial to understanding what can and can't work in this space.

Where we might disagree is on how necessary it is to actually have some of this core infrastructure around rating or labeling the quality of news sources because I believe that's sort of necessary to do many of the other types of work that are crucially important. Understandably, this is very difficult, and I'm not sure that she's as convinced that that is as tractable, in a meaningful way, as I would be.

Julia: Right. Yes. Unsurprisingly, when I've read Danah Boyd's writing on this subject, it's resonated with me so that lines up with your description of it. Great. Well, we'll link ... Is there a particular piece or book or anything by either person that you think we should link to?

Aviv: I think Tim Wu: Is the First Amendment Obsolete, definitely a provocative and interesting read, as is a Danah Boyd's What Hath We Wrought.

Aviv: Okay, great. All right. Well, we'll link to both of those on the podcast website, as well as to your site. Aviv, thank you so much for coming on the show. It's been a pleasure having you.

Julia: Thank you.

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