

Rationally Speaking #192: Jesse Singal on “The problems with implicit bias tests”

Julia: Welcome to Rationally Speaking, the podcast where we explore the borderlands between reason and nonsense. I'm your host, Julia Galef and my guest today is Jesse Singal. Jesse is a journalist, formerly a Senior Editor for New York Magazine's website where he ran the blog, *The Science of Us*, and is now a contributing writer for New York Magazine.

I started following Jesse on Twitter because from time to time these articles would pop up in my newsfeed on Facebook or on Twitter and I'd be like, "Huh. This is an unusually thoughtful, statistically literate example of science journalism." And then I would check the byline and it would be by Jesse, every time. I was like: I should really follow this person on Twitter.

We're going to be talking today about the Implicit Association Test, the most famous test, I would say, of unconscious bias. Particularly racial or gender bias. You may have heard of the IAT, not necessarily by that name but you may have heard it referenced. Because in all of these conversations that our society has been having about sexism in the tech world, or about racism among police officers, in the last few years... people often reference the IAT as evidence that, "Look, even if you don't feel consciously biased in your views of women or minorities, that doesn't mean you aren't. And in fact, most people have these unconscious biases against these groups, which the IAT can reveal."

It's been in the public discourse a lot, probably more than almost any other social psychology instrument. However, as researchers and journalists have been scrutinizing the IAT a little more closely, and the evidence that it's measuring something real and meaningful -- well, it looks like these results might not be as solid as they seemed.

Which is where Jesse comes in. Jesse wrote a piece for New York Magazine that was one of my favorite of his articles, going really in depth into the evidence for the IAT, some of the problems with that evidence, and the interpretations of it in popular discourse. That's what we're going to talk about today.

Jesse, welcome to the show.

Jesse: Thank you for having me. I disagree that I am statistically literate, but I appreciate the compliment.

Julia: Well you do a convincing impression of statistical literacy.

Jesse: Thank you very much.

Julia: The closer that gets to being convincing, the less distinguishable it is from actually being statistically literate.

Jesse: Yeah, exactly.

Julia: Jesse, let's start at the beginning. Can you describe what the Implicit Association Test is? You sit down to take it -- what are you doing? How does it work?

Jesse: In the most basic version of the IAT, you sit down at a computer and you're shown a series of images and words that flash quickly. Some of the words will be negative, some will be positive. You might have, like -- "illness" is a negative word. "Happiness" is a positive word. Those are interspersed with images of black faces and white faces.

This is one version of the test and the most important version of the test. You're basically asked: if you see a good word or a white face, hit the letter E on your keyboard. If you see a bad word or a black face, hit the letter I. Those two combinations are flipped at a certain point.

What's going on under the hood, is the computer is tracking how easily you connect in your brain -- this is the thought at least -- good concepts with white faces, versus good concepts with black faces. Or bad concepts with white faces versus bad concepts with black faces.

The idea is if it takes you longer to connect black faces with good concepts, that means your brain is struggling more and trying to overcome implicit thoughts. According to the test, that is a sign that you are implicitly biased in a way that would favor white people over black people.

Of course the test will also reveal if you have an unconscious preference for black people over white people. But the data they've collected suggests most Americans have an unconscious preference for white faces over black faces. In other versions of the test it might be white sounding names over black sounding names.

Julia: What have the results generally shown? How do people tend to perform on the test?

Jesse: The average American is implicitly biased against black people. But even just describing these results, you need to be very careful -- because as we'll get into, the test-

Julia: Yeah, we'll definitely get into that.

Jesse: That is the lay person -- or what the components of the test would say is that the average American is implicitly biased against black people. And that is more common among white test takers than black test takers. But a solid minority of even black test takers are also implicitly biased against their own race.

Julia: Let's start by digging into this interpretation of the IAT. Reading your article, and some other things I've been reading after your article, has been really helpful. But I have to admit that when I first heard that the IAT might not actually be evidence of bias, my thought was: "Yeah, I agree that just because someone gets this score on the test, that isn't identical to them actually being racist, it's just their score on a test. It's not proof."

But it was really pretty hard for me to come up with a plausible story in which someone would have this unconscious association with white faces and good concepts, or between black faces and bad concepts and not actually be biased.

Maybe you could talk about other ways to interpret the results?

Jesse: The most common alternative explanation, and this has been proposed in literature a lot by several different researchers is: if you're *aware* of negative stereotypes about black people, you might be quicker to associate words like "crime" with black faces or victim or violence with black faces. If that's the case then that could generate a biased score of IAT, even though you're not implicitly biased against black people, you're just aware of these negative stereotypes.

There was one particularly ingenious experiment in which researchers actually created a new non-existent you could call either a race or a species called Noffians, N-O-F-F-I-A-N-S. By inducing in the experiments participants the idea that Noffians are this down-trodden group that society doesn't like, they were able to generate-

Julia: Just reading them paragraphs about Noffians being bad?

Jesse: Yeah. Basically what they did was there are two groups, the Noffians and the Facites, and sometimes they would tell people Noffians are privileged and Facites are oppressed. Other times it would be the reverse. When Noffians were oppressed for example, people would score a higher IAT about Noffians.

Julia: More biased, more -- what we would have naively assumed was bias against Noffians?

Jesse: Yeah, exactly.

Julia: And obviously they can't actually have any reason to think Noffians are bad because they're a made up group. And all we know is they have been oppressed.

Jesse: Right, exactly. That was what was so ingenious about the experiment, was in this case there was no other explanation but that if you view a group as down-trodden, it might boost your IAT score regarding that group. That test doesn't-

Julia: I have a theory -- it's convoluted, but you could have a story in which you have some belief that oppressed groups are oppressed for a good reason. If you know a group is oppressed, then there must be something bad about them.

Jesse: Yeah, definitely. We're talking about, in some cases, a difference in reaction time of a couple hundred milliseconds. If you have a pie that is 200 milliseconds, how big a slice of the pie is actual implicit bias? How big is associations? How big is problems with the way the test is designed, or with error?

The basic problem is that people have assumed the whole pie, or most of it, is something that we can genuinely call implicit bias against a group. But the studies trying to connect IAT scores to actual behavior in a lab setting simply haven't really shown that.

Julia: How do they try to connect this to behavior that we would call biased behavior?

Jesse: There have been, by now, a lot of studies. Where you basically give people the IAT and -- it sounds weird, but you put them in a lab and you give them the opportunity to be a little bit discriminatory.

That can take a lot of different forms. One of them is you give people a chance to interact with a white or a black experiment confederate. You might find that people with high IAT scores are rudier to someone who's black than someone who's white. As judged by a third party observer.

Julia: A third party observer who doesn't know their score on the IAT?

Jesse: Yeah. Third parties are blinded to the whole experiment.

The test first came out in '98. The first study showing this sort of results, showing that it was linked to behavior, came out in 2001. For a decade and a half there were all these other studies that appeared to show this link between IAT scores and behavior.

The test's two creators and main proponents -- Mahzarin Banaji, she's the head of Psychology at Harvard, and Anthony Greenwald, who's another big name in social psychology -- they said, "Look. These are incredible results that show this test can predict behavior better than just about anything else, including situations where you explicitly ask people about their levels of racism."

What researchers have found is their claim doesn't hold up. When you actually take all these studies together and do a rigorous meta-analysis, IAT scores really only explain a tiny chunk of the variation in how racist people act in these experiments.

Julia: Do people's results with the IAT correlate with their explicit self-reported bias? If you just ask them questions about their beliefs about black and white people. Does that connect to their score on the IATs?

Jesse: My understanding is that there is, I think a fairly weak correlation there. By now, they've correlated IAT scores with things like political party, for example. I think there's a fairly strong connection where the higher you score on the IAT, the more likely you are to be politically conservative.

Julia: That seems like pretty indirect evidence for what we're measuring, though. The thing the IAT claims to be measuring.

Jesse: There are so many potential confounds here. Part of the problem is a lot of these experiments were not -- at least according to Hart Blanton, who is one of the smartest critics of the IAT literature -- he doesn't think even the experiments that do show a correlation were designed in a sufficiently rigorous way. That's a view that was echoed by a couple Scandinavian researchers who looked at the evidence and basically said, whatever correlation we find in the landscape of IAT behavior literature, it's so statistically weak, we really can't assume anything to be true here.

Julia: Probably we should separate these two distinct questions about whether the IAT is a good test that actually demonstrates implicit bias. The first question is, were these studies well conducted? If you were trying to replicate this exact phenomenon of people showing a stronger association between white faces and good concepts than between black faces and good concepts, would those studies replicate? As opposed to just being noise or a poorly designed study, et cetera.

But then second -- let's say that the answer to the first question is "Yes, they were well designed studies, a real phenomenon." Then the question is: what does that tell us? Does that have any relation to the things we actually care about? Like people making biased choices or behaving in biased ways.

I guess these two different questions map onto internal validity and external validity. Does that sound right to you? That those are the two questions?

Jesse: Yeah, I think those are the questions. I think where we're at now is that there are consistent patterns in the IAT data where, for example, white people score higher than black people on the black/white IAT. Meaning, if they score higher, their score is more biased.

Julia: So that replicated.

Jesse: Yep. My sense is that that stuff is fairly solid, in that even the critics will say, "Okay. This is a genuine pattern. This is worth investigating."

What has not held up is the external validity. We should go back in a minute and talk about how there's other internal validity issues... But in terms of the

basic external validity question of, does the IAT predict behavior meaningfully? At this point, the answer really appears to be, no.

To the point where even Brian Nosek, who's a big reproducibility guy and was involved in a lot of this research, his name is on a paper -- I'm not sure if it's out yet, or if it's in press -- but the most recent sophisticated meta-analysis showed that IAT scores account for less than 1% of the variance in racist behavior in a lab setting.

In another situation, if someone came to you with this fancy new instrument and said, "Check this out. This is really important. We should spend millions of dollars on it." And you said, "Okay. How good a job does it do explaining what we're trying to measure?" And they said, "Uh. Under 1%" ...There's just no situation where you would find that impressive.

To me, the external validity question really doesn't look good at this point. Or the answer to that question doesn't look good.

Julia: I did indeed want to talk a bit about the internal validity problems too. Why don't you tell us what some of the problems are with measuring the pattern?

Jesse: You do get these consistent patterns, at the big zoomed out level. Especially in terms of how white people versus black people perform on the test. But the test is quite poor when it comes to what's known as test/re-test reliability.

Julia: Like if the same person takes the test multiple times, do they get similar scores?

Jesse: Yeah. Their score will jump around quite a bit. I think if they're white, it's likely that most of those scores will be positive. Meaning, in theory, it indicates they're biased. My article has the nitty gritty statistical stuff. But basically, this test does not come close to the level of test / re-test reliability we would expect for any instrument professionally used to measure anxiety, depression or anything else.

The reason why it blew up the way it did and became this viral sensation, despite performing so poorly in terms of its psychometric attributes, is an important question. I wish we had a better sense of why that is. But it just doesn't come close to the level of performance that one would expect.

Julia: Thinking more about how it could be the case that the pattern is real but it doesn't actually translate into biased behavior... Someone, I think it was a philosopher, but I'm blanking on his name now, said: it seems to them what matters isn't whether people have an association between, say, the concept of black people and the concept of "bad" or "dangerous." What matters is how *central* the concept dangerous is to their conception of blackness.

Because if it's not central, then it can easily be swamped by other contextual associations. I think the example they gave was: if dangerousness is central to your concept of blackness, then you're going to associate that with a black person no matter what context you view them in. Like, whether you see a black person in a church or a university or a street corner.

But if dangerousness is just very secondary or peripheral to your concept of blackness, then *maybe* you'll associate a black person with dangerousness in some context like a street corner. But as soon as the context switches to say a church or a university, there's no longer that association for you at all.

The IAT, it's this very abstract, contrived -- it's really contextless. It doesn't really measure the centrality or the context dependence of the associations that it's testing for. What do you think about that?

Jesse: Yeah, I mean that makes sense to me. I think what's striking is the way that even though all we're measuring is this very slight fraction of a second, in many cases, reaction time difference... The folks who created this test and then helped publicize it, and wrote a big book about it, simply stated -- in my opinion without evidence -- that this test could explain everything from police shootings to problems in terms of who gets to rent what home. They basically said, this test can explain a big chunk of racism in the U.S.

I just don't think that claim is merited by the evidence at all. I think all the focus on the IAT has potentially sucked a little bit of the oxygen out of the room, and diverted funding and attention from other better or more rigorous ways of understanding race from a social psychological perspective.

Julia: Yeah. One really important point that I want to make sure comes across in this conversation is that -- so, let's just assume that the IAT is completely bullshit. That would not imply that there's no such thing as implicit bias, right? It just means this one test that was designed to measure implicit bias is not a good test.

In fact, I, personally would be shocked if there is no such thing as implicit bias. All of my priors and a bunch of anecdotal evidence suggests to me that there is. And by priors I mean, we know that a ton of cognition is unconscious. We know we're influenced by patterns that we're unconsciously picking up, or that we think we're picking up -- our brain is picking up -- without being aware of it. And there has been plenty of stereotyping by race or gender in our culture historically, so I would expect that to influence our unconscious pattern-matching algorithms.

Maybe the fact that it feels like such a common sense thing to exist is why people have sort of over-played the evidence for this test. Because on some level they're like, "Oh, well this should make so much sense, it must be true," and they don't worry too much about the fine details of the evidence.

Jesse: Yeah, and well, you're touching on one of the frustrations about trying to write about this is a rigorous way, which is -- when you point out the weaknesses of the

test, you sort of get two annoying responses, one from each side of the political aisle, to over simplify a little bit.

Julia: It's like the story of your life on Twitter, Jesse.

Jesse: Yeah, you could say that.

Julia: I'm really impressed by the equal amounts of vitriol you get from both sides.

Jesse: Yeah, which I find weird because I really do consider myself pretty far to the left, at least by American standards.

Julia: Yeah, I also would consider yourself on the ... Yeah, in an American context you seem pretty clearly on the left to me.

Jesse: Yeah, but you know, you read an article like this and people on the left will say, "Oh, so you're saying implicit bias isn't real? You're saying racism isn't an important thing to study," which no, I'm not saying that.

People on the right will say, "Well, this clearly shows that the focus on implicit bias is misplaced because this shows implicit bias isn't real," when in fact, as you just said, this is just one instrument. If I gave you a thermometer, and you show that the thermometer doesn't do a good job measuring temperature, you would not infer from that that temperature isn't an important thing to measure.

The evidence for implicit bias -- I think there is a fair amount of solid, empirical evidence for it. When you look at, for example, studies where they send out a bunch of job applications with stereotypically black versus white names, there is this consistent pattern -- where if you have a white name, having a white name grants you sort of a bonus in the probability you'll be granted an interview. That, to me... sending resumes out in the 21st Century when we know people claim not to have explicit bias, that to me is fairly solid evidence for implicit bias.

Julia: It's also a little more, I mean I know it's a test and not ... Well no, no, sorry. It's actually an in-the-field, real test. People think they're actually deciding who to hire, right?

Jesse: Yeah.

Julia: Right, that seems like much more direct evidence of what we care about, than the IAT.

Jesse: I would say so. People should look up the work of ... I took a class with her. Her name is Devah Page or D-E-V-A-H. She's done some of these resume studies, as have a lot of other academics.

And yeah, I mean you can tell some sort of pretzel-like story where, "Oh, that's actually explicit bias but they're hiding it and no one can observe them..." But to me, as you said, implicit bias -- everything we know about sort of system one versus

system two processing, and the way humans really carve people up into categories. I absolutely think implicit bias is real. In some settings it probably matters a lot.

I also think what's happened over the last two decades in social psychology is that this one very faulty test has convinced people that implicit bias is probably the thing to study to understand racial discrepancies. When in reality there's all sorts of deeply rooted structural stuff going on, that doesn't really rely on implicit bias to reproduce these inequalities. To use the sort of lefty sociological language.

Julia: Right. I guess, now that we're talking about these more meaningful and fruitful tests of implicit bias, like the resume studies, I guess I'm wondering what is the value add of an IAT?

Okay, if we want to test people for implicit bias, why don't we just give them resume tests, for example? I know that we couldn't just give them a single resume and infer their level of implicit bias from that, because the bias there would be measured over a set of resumes. You have to look at whether on average they show bias against black named resumes. But you could do that. You could just give people twenty resumes to evaluate, some of which have black names and some of which have white names, and you would randomize the name, and see how they evaluate them. If they show a correlation.

Or you could give them a pretend court case with a defendant who's either black or white, and ask: do you think this person's guilty, or what kind of sentence do you think they should get? And over twenty of these you could then test, you could then see if people are harsher, less forgiving of black defendants than white defendants.

Why aren't we just doing that? That's so much more direct.

Jesse: I think, well, for one thing, the IAT -- you can sit down and take it in about ten minutes. It's very simple and straightforward to take.

I also think that the proponents of the test would claim it's measuring something a bit more primal, on like a really basic cognitive level, that isn't captured by the more intellectual system two process of sitting down and evaluating resumes and evidence.

Julia: I see.

Jesse: This is where stuff gets really complicated. Because if you talk about a hiring manager looking carefully at different resumes, how much are they influenced by system one thinking versus system two thinking? How can you even answer that question empirically? The nice thing about the IAT -- if you think it works -- is it really boils things down to gut impulse, system one thinking and it doesn't let other factors creep in. Again, in theory.

Julia: All right, so let's say it's true that to some extent implicit associations predict bias. I hear you that the evidence is pretty shaky on that, but let's just assume it were true.

Does that mean, or would that mean that if we wanted to get people to act in a less biased way, that the right approach would be to change those associations? Like, get them to start re-associating black and white people so they are more likely to associate black people with “good” instead of bad, and that would cause them then to behave in a less bias way?

What do you think of that connection? It's like jumping from using this as prediction to assuming a cause --

Jesse: Intervention.

Julia: Yeah, exactly.

Jesse: Got you. I would say – again, we're now in a fantasy world where the test does predict something meaningful.

There's still no evidence that you can actually attack these associations and change them. There's now been -- I'm less versed in this research, but there have now been a number of studies attempting to build interventions that basically do what you just said, which is sort of try to re-wire people's associations. And that big meta analysis I mentioned that Nozick is a co-author on, the main thing it was looking at was the effectiveness of these interventions.

And I have a quote in my article here, from Calvin Lai, who's this Harvard -- I think he's a post-doc, who worked on that study. He basically said, "We found no reason to believe that these interventions work, that they do what they're advertised to do." And I don't find that surprising, because the test isn't really measuring anything in the first place. Even if you assume it is, there's still no evidence that you should, that this is the most useful use of inevitably limited resources to try to fight racism.

Julia: Are there versions of the IAT on topics other than race, that hold up better? That are more stable, have more ... What did you call it, test, re-test, you know where people's score stays basically stable no matter how many times they take the test?

Jesse: Yeah, there ... Oh sorry.

Julia: Well, I was just going to ask if there are versions of the IAT that score better on that metric and are also more predictive of real world behavior. What about IAT for gender instead of race?

Jesse: Right. The gender one is, if anything, weirder. Because it fairly consistently finds that women are more implicitly biased against women than men are.

I'm trying to pull up an article that I wrote about this... but it just gives all these weird results that we're supposed to take at face value. I'm looking at this chart and it basically shows that white women have the highest level of implicit bias. And any kind of man, white men, black men, Hispanic men are less implicitly biased when it comes to gender.

Julia: Meaning that they're less likely, they seem to have less of an implicit association between female faces or names, with concepts like "good"? Or what is the actual association they were testing for?

Jesse: Yeah, this is basically, let me see which one this is. This is the compilation of everything they ask on all the tests that were taken on Project Implicit. I'd have to actually go back and see how they frame this.

It's basically just reported as implicit gender bias, which we can take safely to mean just that men are better than women, or better qualified for certain positions than women.

In this view, it varies a little bit depending on how liberal or conservative you are, but generally speaking all women are more biased than all men, more implicitly biased.

Julia: Do you think men show any bias, or is it just not distinguishable from zero?

Jesse: They show some bias, yeah. It looks like a moderate amount of bias.

So what the IAT asks us to do is accept this version of reality where women are just consistently and significantly more implicitly biased against women than men are, which... implicit bias is this sort of fuzzy and wooly concept where you *could* have bias against yourself. You could have all these weird societal factors... but I'm just not sure why we should accept this idea that women are significantly, across the political spectrum, more biased against women than men.

What's funny, I'm looking at this now -- even strongly liberal women, they're more implicitly biased against women than far, far right men, so think about that for a minute.

Julia: Whoa, wait a minute. I was sort of skeptical when you were saying that women showed more bias against women than men, because I was like, "Well, you know... If there's a culturally dominant stereotype about what women are good at or bad at, I wouldn't necessarily expect women to buy into that less than men." It wasn't shocking to me that women would show more of those biases.

But the strong liberal versus strong conservative -- that just goes against all of my common sense about gender bias.

Jesse: Yeah, so this is what I'm sort of saying. When I wrote this article, this is a second article, I was responding to a 538 piece by a very talented reporter, who reported these associations at face value. And I don't think he was skeptical enough.

To me, these tests are aren't magical. Researchers aren't infallible. You're telling me that a far left woman is more implicitly biased against women than Jerry Farwell, Jr.

Julia: Right.

Jesse: Maybe I'd be more likely to believe it if this test had shown that it was ready for prime time in terms of all the psychometric properties, but I just find that to be a very hard result to believe.

Julia: Yeah. This almost makes me wonder, how easy is the IAT to game? Could it be the case, for example, that far right men are like, "Oh these liberal researchers are trying to prove that people are sexist. I'll show them, and I'll associate women with all the career-minded words," or whatever.

Jesse: Right. One of the big selling points of the test, especially I think in its earlier days, was that you can't really game it. It's sort of this computer algorithm truth serum that reveals your deepest beliefs.

Julia: I really do see why this became so beloved by the public. It's such a sexy idea for a test.

Jesse: Oh, it's so compelling. Yeah, I want to circle back to some of those sort of political ramifications for liberals, but the test proponents have not told a consistent story over whether or not these are impulses we can control or whether or not it can be gamed. You can see that a little bit.

I'm not sure these two ideas are entirely contradictory, maybe you tell me. On one hand you have the idea that this test is revealing something about ourselves we can't control. It can't be gamed. Then the other hand, you have, "But there are these interventions that can reduce it, that can fix it." I guess you could sort of see those two ideas ...

Julia: Those don't seem contradictory to me.

Jesse: Got you, so you're saying maybe if like you make a conscious effort to fix them then you can?

Julia: Yeah. You can reconcile them with each other by just saying: Taking this test, you can't hide your current level of bias. That's going to come out. But that doesn't mean you can't change your level of bias over time.

Jesse: Got you. Yeah, okay. That part's not that contradictory.

Julia: That's the way I look at it anyway.

Jesse: I guess where they have been contradictory is different researchers, at different times, have made different claims over to what extent, in one test taking session, you can control your results or game it.

Julia: Yeah, that makes sense. Great, so let's talk about -- you were alluding to the political implications of these results, for liberals in particular?

Jesse: Yeah, so I'm basically working on a book about shoddy ideas in social science and why they catch on, and what they can tell us about society. My IAT chapter is going

to be in some regards similar to the article I wrote, just talking about the method, the logical issues.

But I'm also interested in the way that -- in a way that might sound strange, I think this test tells white liberals a story they want to hear, and I'm including myself in that because I'm very much white and very much a liberal. There's something about the experience of taking this test that can make white liberals feel like they are taking part in the fight against racism. Despite -- I'll try to be gentle, but often times white liberals don't really do much but take the IAT or tweet about racial injustice. Oftentimes, because of the crappy system that we have set up, we are perpetuating many of these forms of inequality.

I think the IAT might give us an easy way out, to feel like we're fighting the good fight without actually doing anything.

Julia: I really appreciated this part in your article, where you were talking -- I think it was your article, where you were talking about the kind of performative nature of people reporting their scores on the IAT.

Jesse: Yeah.

Julia: There's this sort of ritual that people seem to go through of reporting their score with this really troubled, emotional tone, almost like a confessional, like ... This wasn't your phrasing, but it reminded me of some of the circles in the early '70s, like at Esalen or something, where people were sort of confessing their sins to the group in this really emotional way, almost exercising their bigoted demons or something.

Jesse: Yeah. There's this sort of ... For some people who take the test, not just white people but mostly white people, for understandable reasons, there is this sense of, "This sin within me has been discovered, and I can expurgate it by telling people about it, and by telling people about how I respond to this test emotionally."

It's just striking, because the test doesn't really do anything, as I would argue we now know. It's just another way, in my view, for white liberals to talk about caring about this stuff without actually putting their money where their mouth is. Which brings us to the ... Yeah?

Julia: Well, I want to stick up a little bit for the white liberals.

Jesse: Sure.

Julia: If I imagine this test actually measuring a real thing, and being reliable and so on, then it does seem ... It's not the most high-effort or costly thing you can do, but it still seems pretty valuable for people to post their results and say, "Huh, turns out actually I am biased, too," because it helps with the story of "Racism is real, and just because you think you're not a racist, doesn't mean you're not acting in a way that hurts minorities or something."

That does seem ... Well, I could also tell the opposite story. This is something that I talked about in my previous episode with Seth Stephens-Davidowitz, which is that it's not clear that making it common knowledge that everyone is racist is actually helpful. Certainly, it seems plausible it could be. But you could also tell this other story where it just normalizes racism, and then everyone's like, "Oh, well, I guess if everyone's racist, then it's fine for me to be racist," and it makes it seem more acceptable. I don't actually know.

But at least, I can tell a plausible story in which this would be a valuable thing for white liberals to do in a world in which the IAT was measuring something real and meaningful. That's a lot of qualifications, but.

Jesse: Sure. Yeah. No, no. I appreciate your impulse, because I do think overall, anything you can do to interrogate your own role in propping up racism or propping up any bad institution is good. I just think A, the performative aspect of this, and B, the extent to which it just isn't anchored to anything in the real world -- which, again, is something the average test-taker doesn't know -- it just seems to me it just doesn't get us anywhere. It's just more white people talking about how dismayed they are at their own racism without then ...

I just don't find IAT conversations tend to lead to discussions about what to do. I think it leaves people maybe feeling a little bit helpless. I think from a ... I'm not an activist, but I think it's safe to say, psychologically, if you make people feel helpless or mad at themselves, that isn't necessarily the best way to then get them to actually do something about those feelings.

Julia: Yeah. It also, and this is another good point you raised in the article ... It also strikes me as kind of unethical to keep giving people this test that tells them that they're biased, without necessarily good evidence for that. Because as you say, it is, it can be psychologically harmful. It's an unpleasant thing to have to believe about yourself. It's ironic, because the IRB, the, what's it called, Institutional Review Board?

Jesse: Yeah.

Julia: The board that's supposed to prevent unethical studies from being done and make sure that participants are being treated well, and they have informed consent, and they're debriefed, and all these things. In so many cases, the IRB just dramatically overreaches and bogs down completely harmless studies in months of red tape. That's maddening.

Yet, at the same time, stuff like this happens that actually does seem harmful and somehow slip through the cracks. I don't really understand how this works.

Jesse: Yeah. Did you see the Scott Alexander piece?

Julia: Yes. That's what I'm thinking of. We should link to that.

Jesse: Yeah. Slate Star Codex is his blog. He has a great piece about navigating the incredibly byzantine IRB process. Yeah.

This is another point I hope I can go deeper into in the chapter. Brian Nosek, who I quite have a lot of respect for, when I asked him about this, he said, "If my 5th grade daughter gets a bad grade on a math test, it's just one piece of information she should fit into her broader understanding of how good she is at math." He compared that to the IAT.

I don't think that comparison works, because since the test was created, so many people who have taken it, including the co-founders, talked about the searing nature of the emotional experience of getting your test result and finding out that you're biased. So it's very clear that taking this test has a big emotional impact on some people.

What's another situation in which we would be okay with a psychological test being promoted by Harvard eliciting in people this sort of reaction, despite the fact that it isn't clear what exactly if anything it's measuring? If Harvard was offering an anxiety or depression test that had these psychometric properties, and that caused people to get really upset, there would be rightfully some serious ethical questions about that test.

I don't know. Summing it all up, racism is just such a tricky, emotional subject, and we so desperately want easy answers about how to address it, that I think our desire for those answers swamps some of our other more thoughtful impulses, maybe.

Julia: Defenders of the scientific process, which frequently includes me, will often say, "Okay, look. Yes, we were wrong about this phenomenon before. We've had to revise our models of it. Some of these results were wrong and they got corrected. But that's not a flaw in science. That's how science is supposed to work."

Of course, on the other hand, some errors are just part of doing science, but other errors are more a result of sloppiness, or intellectual fudging, or even intellectual dishonesty. And that's actually a bug and not a feature in the scientific process.

I'm curious whether in your opinion the huge excitement over the IAT, and then having to walk a ton of that back, whether that is a case of science functioning healthfully and self-correcting? Or was there a flaw in how we went about -- "We," you know, "science" -- went about studying and promoting the IAT?

Jesse: "We" meaning me and you, over the last 20 years...

Julia: Right.

Jesse: I think I have complicated views on that. But I think there is no way to look at this whole thing and not consider it at least a minor disaster. For 20 years, you've had a huge amount of resources flowing toward a test that is just so psychometrically weak, and to me, really misleading to a lot of people. That, that is a failure of science. In terms of how to divvy up the blame or who should have done what ...

Julia: Where should I point my finger?

Jesse: Exactly. I blame white liberals!

... I think folks like Greenwald and Banaji, who really were the cheerleaders for this test -- I think they're trying to be good scientists. I think they take these issues seriously. I also think they made claims that significantly ran ahead of the evidence. That's most striking in some of the text of their book, where they really say, "This test does a great job predicting behavior, better than anything else we have, better than explicit measures."

Two years later, they themselves are admitting in an academic journal that no one's going to read, "This test should not be used to diagnose individual levels of implicit bias." They themselves have admitted that.

Why was there all this excitement over a test that, 20 years in, the architects of the test had to say, "You shouldn't use this on individuals." If you can't use it on individuals ... Really? We're going to now use it to make these big maps of racism in America, as some people have done? We're going to use it to compare big populations?

I don't know. I'm torn. I don't want to be too hard on them. I just think every step of the way, they got excited, and they saw the attention it was getting, and I think they made the mistake a lot of scientists make, which is they sensed they had a really exciting idea on their hand, and then let the excitement get the best of them.

Julia: Yeah. There's also this thing that I've seen some of ... Not from all proponents of the IAT, but from some of them, that's like: when they encounter criticism or pushback on the claims, they use the defense, "If you're criticizing the IAT, then maybe it's because you're a racist, or you don't care about racism," or something like that.

Jesse: Yeah.

Julia: That is such bad form. That's so toxic to being able to talk about evidence and converge on the truth in the long run. I have very little sympathy for that.

Jesse: Yeah. Look, I put it in the article -- it's not a secret, but Banaji, who is a head of psych at Harvard... she said in emails, like, she basically made that argument. That people who criticize the IAT want to go back to a world of

segregation. She didn't say that exactly, but she basically, she strongly implied they're racist.

Julia: Or that that's the plausible motivation for why someone would doubt the IAT.

Jesse: Yeah. Which, if you're going to make that claim -- what's the point of debating anything, if you're just going to immediately jump to that? There are obviously a lot of people who are motivated by racism, but in this case, she's talking about a lot of researchers publishing in top journals. About a test that she herself in writing has said is too weak to tell individuals how implicitly biased they are, even as Harvard University continues to do exactly just that every day.

I think she should be in a bit more of a position of defending why she's made these statements about her test, and not accusing what I see to be good faith critics of it of being racist or whatever.

Julia: Yeah. There's also ... I was trying to be charitable to "Science is not going to be perfect on its first try with every investigation it does, or every conclusion it reaches." And I still stand behind that.

But the more I think about the discourse around this test, both early on and now that there's been a lot more questioning of it, the discourse seems really bad. Even -- this is less toxic than accusing your critics of racism, but even the response of, "Well, maybe the IAT is not perfect, or maybe it doesn't actually show conclusively what it said it did, but it's still educational. It's still helping teach people about racism."

Jesse: Right.

Julia: That's such a weak ...

Jesse: Yeah.

Julia: That's such a rationalization. It's not ... Anyway, sorry. Go on.

Jesse: Well, no. That's exactly right. Again, all you have to do to understand the weakness of these arguments is swap out any other thing that social psychologists or any psychologist might want to measure. "This test won't actually tell you how suicidal you are, but it's a good opportunity to teach you about suicide..."

Julia: Right. Right. "It gives you a score, a suicidal score, but that suicidal score isn't meant as a diagnosis! And anyway, the whole point was just to educate people about suicide!"

Jesse: Right. What drives me crazy, and you see this in so many different areas ...

Julia: "And If you oppose our test, then you think people should commit suicide."

Jesse: Exactly. Right. Again, there are certain concepts that seize our brain and turn off our usual ways of thinking and our usual standards. I do think, because racism is so pervasive and offers up so many salient examples of how horribly people in this country are treated for no reason -- but that shouldn't short circuit us and cause us to make arguments that we wouldn't make in other cases.

The other one I hear a lot is, "Sure. This test doesn't really predict anything, but it's better than anything else we have." Okay. If I have a thermometer that is consistently off by 50 degrees Fahrenheit, that's a little bit better than random, but would I ever use that for professional purposes? Would I consider its readings to be accurate? Of course not.

It's just weird to see a totally different set of standards apply to the IAT than would apply to basically any other scientific project.

Julia: Right. All right. Well, we're just about out of time.

Jesse: Did we ... I think we've fixed racism in the last ...

Julia: Yeah. That's right. It took us less than an hour! I don't know what the rest of society was doing with their time the last hundred years. Very efficient.

Before we close, Jesse, I want to give you the opportunity to recommend the Rationally Speaking pick of the episode. This is a book, or a blog post, or article, something that has influenced your thinking in some way. What would your pick be?

Jesse: Everyone should read *Galileo's Middle Finger*, by Alice Dreger.

Julia: Oh, that's also on my reading list.

Jesse: Yeah. It's so good. It launched me on this whole ... This is another podcast for another day, but I've written a little bit about gender dysphoria, and especially the question of what you do when a five-year-old has gender dysphoria.

Julia: Gender dysphoria is identifying with a gender that doesn't match their biological sex?

Jesse: Yeah... sometimes --

Julia: Sorry, did I just open up a huge can of worms in trying to define gender dysphoria?

Jesse: It's okay. We'll resolve this on Twitter.

No, basically, it just means basically a feeling of discomfort with your own sex in one way or another. And some people interpret that as, "I'm the other sex." Some interpret it as, "I have no gender." It gets complicated.

Dreger's book really opened my eyes to the way good, Leftie, social justice can sometimes be at loggerheads with good, rigorous science. She offers a number of case studies, including one about transgender issues. She just has this way that I really admire of taking these complicated academic fights and turning them almost into mysteries you want to solve, and these really compelling narratives. In a way no other book has done, her book just sort of nudged my writing career in a different direction, and got me involved in stuff I didn't think I'd get involved with. I just think it's a wonderful book everyone should read.

Julia: Fantastic. Well, I'm grateful to her for doing that, although, I feel bad for your blood pressure as you wade into the swamp on Twitter every day and get attacked from both sides. Thanks for fighting the good fight.

Jesse: That's okay. It's the price of admission. I should say, a lot of the times, people who are mad at me on Twitter, they have legitimate grievances, yeah, and that people have their own stuff going on, so I try not to take Twitter too, too seriously.

Julia: Yeah. Well, easier said than done, but something to strive for.

Jesse: Yeah.

Julia: All right. Well, we'll link to Dreger's book, as well as to your article. We should also link to Scott's post about the IRB, which everyone should read.

Jesse: Definitely.

Julia: Jesse, thanks so much for being on the show. It's been a pleasure talking to you.

Jesse: Thanks, Julia. This was a lot of fun.

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