Opinion Science Podcast

Hosted by Andy Luttrell

"Unconscious" Bias? with Adam Hahn

April 11th, 2022

Web: http://opinionsciencepodcast.com/
Twitter: @OpinionSciPod
Facebook: OpinionSciPod

<< News clips about unconscious bias. >>

Andy Luttrell:
The idea that our minds unknowingly harbor prejudiced views about race, gender, and so on, is a captivating concept. Like there are thoughts running through my mind that I can't know? The clips you just heard highlight how news and education programs often convey exactly this message. And to let them off the hook, what they're saying echoes messages coming straight out of social psychology, which has been studying implicit attitudes for decades. Even the textbook I use when I teach this subject defines implicit bias as "reactions toward groups or individuals that occur automatically and outside conscious awareness."

And listen, there may very well be opinions we hold unconsciously—maybe this implicit association test does capture them. But this part has kind of been taken as a given for a long time. It's time we take a step back to really question whether the stuff we call "implicit" is really "unconscious bias"

First, a very, very quick introduction to where some of these ideas came from. We can look back to the 1980s when social psychologists like Russ Fazio and John Bargh were finding that whether we like or dislike something seems to pop into people's minds automatically when they encounter it. I look at my morning coffee: I like it! I see a skunk in my back yard: I don't like it! (Sorry, skunks.) I see my daughter's face after works: I like it! And it's not just that they had the idea that these thoughts occur automatically, they devised a method called "sequential priming" to capture these automatic thoughts.

Okay, so in the 1990s, two other social psychologists—Mahzarin Banaji and Tony Greenwald—were inspired by some ideas from research on human memory and presented what they called "implicit social cognition." And they were very clear that they "use the term 'implicit' to refer to those processes that operate without the actor's conscious awareness." And they also developed a method to capture these so-called "implicit evaluations." It's called the "Implicit Association Test" or "IAT," and it's become a very popular procedure in social science research. Essentially the test is like a little game—it's no Tetris, but I think of it as a game nonetheless. You see a bunch of stuff pop up on a screen and your job is to categorize those things in different ways. It depends on what the researchers are trying to measure, but if we wanted to measure implicit racial biases, you might

see pictures of people's faces, which you'd categorize as "Black" or "White," and you'd see words like "beautiful," "ugly," "wonderful," which you'd categorize as "good" or "bad." And you have to do this as fast as you can, but the game is set up in such a way that certain parts of the game are really easy if you strongly associate "Black" with "Bad" and "White" with "Good." Other parts of the game are much harder if that's your bias. So after some behind-the-scenes calculations, researchers can get a reasonably informative indication of people's biases in favor of one group over another.

But okay, the key question here is: are people consciously aware of the biases that the IAT measures? Lots of writing by psychologists and presentations to the public give the idea that yeah, this test tells us something about our unconscious biases.

So, I actually talked to Mahzarin Banaji—one of the IAT's originators—back in 2020, and I asked her about the claim that the IAT uncovers unconscious biases. Here's what she had to say:

Mahzarin Banaji:
If it were the case that these were all within conscious awareness, we wouldn't get the hate mail we do, okay? So, all I can tell you is that that is the signature result…However, I think your question is a very good one, because in the old days, I think we were mistaken. I think we thought these were two separate systems with not much in common, that these were really, just like with memory, they were separate memory systems. That this is separate systems in the brain, that one responded to the conscious probe, the other to the less conscious one, and they didn't speak to each other. And that's largely been shown to be false.

Andy Luttrell:
Go check out Episode 16 of Opinion Science for the full conversation. It's one of my favorites. But we keep the theme going today with a new set of glasses on.

You're listening to Opinion Science, the show about our opinions, where they come from, and how they change. I'm Andy Luttrell. And this week I talk to Adam Hahn, Senior Lecturer in Psychology at the University of Bath. Adam has spent a lot of time thinking about how strong a case there is for calling implicit attitudes "unconscious attitudes" and running careful experiments to see whether people actually are aware of these thoughts and feelings we've been trying to say they can't perceive. Anyhow, I talked to Adam Hahn about what implicit attitudes are, whether people can be aware of them, and what this means for us going forward.

<< Transition >>

Andy Luttrell:
Okay, if implicit attitudes are being defined as unconscious, what does that even mean? Even if it were true, what would it mean that these attitudes are unconscious? And it sounds like that could mean several things, so what could that mean?

Adam Hahn:
Right. So, if you look at the literature in cognitive psychology and social psychology, it's surprisingly ill defined, so we talk about awareness being a feature of automaticity, so when

something is automatic it might happen outside of awareness and that's as far as it goes. Now, the sort of straightforward definition that I claim, or I would claim people hear when you say unconscious racism, and that's something I've been trying to change, is this idea that something is inaccessible. So, an unconscious process is a process that you have but you cannot know. It's impossible for you to observe. And there's a lot of unconscious processing in our brain. That's not esoteric anymore, but we now know that for instance the light that's reflected on the retina of our eyes transforms into a three-dimensional image, but you can't observe any of those processes. All you can say is, "I see you," and that's it. And it's 3D.

So, there's a lot of unconscious stuff going on in our brain that's happening that's impossible for us to access. And I would argue that's what people hear when you tell them, "You're an unconscious racist." Oh, something inside of me is having these… is producing a prejudiced reaction, and I can't know, and that's sort of the danger of that definition. Interestingly, going back to what I said earlier, in the 15 years I've been teaching this because I have some similar problems with my students as you mentioned earlier, so you teach your students about all of this and in the end they say, "Oh, it's unconscious bias." So, what people mean is it's independent of awareness often, so when we use unconscious colloquially, we usually tend to describe a state, not a trait of a cognition. So, any cognition can at some point be inside or outside of awareness, right? Like I'm not consciously aware at all times of the day that my heart is beating, but I could pay attention to it, but I don't always, right?

Our attention is very limited. We can sort of hold at most seven things in our minds at the same time. And so, what people often mean when they say, "Oh, I wasn't aware of that," is not it was impossible for me to know that this happens. Instead, what they mean is it wasn't in the focus of my attention. And when you define it that way, there's a lot of evidence, and this is sort of what I've come to over the years. There's a lot of evidence that most people are really not aware that they're biased. So, we do have unconscious biases in this situational definition. We all have biases that we don't always notice. And I think that's what resonates with people, that you do an IET, to many people it's surprising because they've never thought about these things, and so they come to the conclusion, "Oh, yes. Of course. This is a measure of unconscious bias. It's the measure of a bias that I don't know."

What's so tricky about that definition and how imprecise it is is that I would still argue people hear you have inaccessible bias that you're not responsible for, which is a completely different thing.

Andy Luttrell:
Yeah, so just to think through that a little bit, so this difference between state and trait possibilities, like one version of that would be like these attitudes, these feelings that you hold, they're locked in a box, shut tight in your head, and you will never be able to see them, right? They're there. They live in your head. And sometimes they leak out and they affect the things that you do, but you'll never be able to see them. That's like the state version, whereas the trait version is to say like no, they're in your head, you just are never really looking at them, right? They're the little ghosts in your house and if you could just turn your head at the right moment, you'd go, "Whoa! Look, that's there." But most of the time they're sort of wandering around unchecked, right?

And so, that's just two different ways of thinking about what does it mean that I'm unconscious of these attitudes. One is they're locked away; I'll never know they're there. The other is just like, "Oh, they're there. I could pay attention but I'm usually not." That's what you're saying, right?

Adam Hahn:
That is what I'm saying. Exactly. And the second one is state. I think you said it the other way around, so trait would be you never know… It's a trait of the cognition that it's inaccessible, that it's unconscious, and state would be it may be unconscious at different points in time. So, part of the thing that also draws people to noticing that these things are unconscious in the state sense is that that is a very fundamental difference between whenever we're picking up on implicit measures as opposed to explicit measures, right? Your explicitly stated attitude on how much you like different groups of people is something that is in awareness at all times, so it can be very easily drawn into awareness, so people walk around knowing that. Whereas whatever we're picking up with implicit bias measures is more or less awareness independent, right? You will have them even if you don't want to. They will exist whether or not you pay attention to it.

And I think that also bolsters this feeling that people think we should conceptualize these things as unconscious, because they're always there. And you may not pay attention and they will still be there.

Andy Luttrell:
Yeah. I like the characterization of these kinds of attitudes as inescapable, right? I know that that's been a term that's often levied at these, like they happen, right? You encounter something. This overall impression you have of it comes to mind whether you want it to or not and you're saying, "I could focus my attention on that," but oftentimes we just let these inescapable thoughts happen without really knowing that they're going on. So, like the evidence that's often I think been used to kind of propel forward this idea that these things are unconscious is kind of what you were just saying, right? This difference between what we've called implicit and explicit attitudes, right? And so, you say we measure implicit attitudes by using these kind of sneaky games that sort of catch people's thoughts as they happen, right? We can sort of bottle that up even if people don't realize they're doing it. We can sort of maneuver the game such that we can tell, "Oh, you have a bias in favor of one of these things over another." And explicit measures are just coming out and saying like, "Hey, do you like this thing? Do you like that thing? Do you like this thing more than that thing?"

And for so long, it seemed like oh, when people are asked directly, they give answers that aren't always consistent with the answers we get from these sneaky games. And so, we go, "Ah, well, the conscious thing they did is different from this game that they're playing, therefore the game's gotta be catching something unconscious." And so, I know that you would disagree with that, which is like what are these… How else could we say sure, these explicit things and these implicit things can be different, but it's not because the implicit stuff is outside of our awareness?

Adam Hahn:
Yeah. So, I could explain sort of different theoretical accounts, but I think the most straightforward way as a layperson to understand how it's different is just make yourself look at a bunch of pictures of say women, and a bunch of pictures of men, or a bunch of pictures of Black people, and white

people, and think about what's the first reaction that's being triggered by that. And now ask yourself, "How much do I like Black people, or white people, or men, or women?" So, this idea that the only thing that would matter for the answer for that question was that first affective reaction is not intuitive once you realize that what the IAT is picking up is just your reaction towards the pictures, right? That is one thing and it's very likely largely culturally learned, although it's more complicated than that, but that's one part of it.

And then when I asked you, "How do you feel about these things?" You retrieve all of the knowledge you have in long-term memory about these things. So, you might think about all your friends that belong to a certain category, people you admire, people you look up to. You might also consider sort of what's the appropriate thing to say here. You might feel like, "Oh, the first thought that comes to mind might be this thought, but when I think about it longer this other thought is much more important." And so, then you come to a conclusion this is how I feel about this group, and we call this process a propositional process because you're sort of like solving a syllogism. There's all these different thoughts that are salient at this moment in time and you try to create consistency and decide what does that mean about how I feel? So, there's this initial feeling, but there's also, "I really love Barack Obama. I love Denzell Washington. There's all of these people I admire." Then there's thoughts about what should I say, what should I not say? Then there's all this knowledge I have about horrible things that white people do, but I'm also a white person maybe, so this is an example of a white person making a racial judgment.

So, you have all of these thoughts, and you sort of have to come to a conclusion, and so of course all of this happens in sort of split seconds, but you sort of have to decide. Which pieces of information are more important here? And you're gonna discount some and accept others, and then your attention, going back to that point, is not going to be focused on all of those things at the same time. Maybe in a questionnaire you're only going to think about what am I gonna look like, or maybe you're only gonna think about how much you love Michelle Obama because you just had a conversation about her and how awesome she is. And so, then your answer is going to be on whatever is salient at that point in time when someone's asking you. An IAT is picking up your reaction towards pictures or whatever stimuli it uses, and obviously those are not the same. That's not the same as making a sort of sociocultural political judgment on an entire group of people.

Andy Luttrell:
Yeah. Another example of that that I like that takes it a little outside of the prejudice domain is what happens when people want to quit smoking, right? So, you go, "I have decided. I have learned the evidence. Smoking is bad for me. My friends don't approve of it. I can't go anywhere. It's cold outside and I don't want to have to go outside to do it." But nevertheless, if you've smoked for many years, your first impulse when you see a cigarette might be a positive one. And it's not that you don't know. You go, "Yeah. No, I know. You don't have to tell me that I have this impulsive approach orientation toward this thing." But give me half a second more to think about it and I can tell you like, "No, no, no. I have come to the conclusion that this is not something I want to do."

And so, what you're saying is these implicit measures would have caught you in that moment of going, "Ooh, that sounds good." And your explicit measures are incorporating all this other stuff. But that doesn't have to mean that I don't know that I have that first positive impulse, right? That's what you're saying too, right?

Adam Hahn:
Yeah. That's exactly it. So, there could be a lot of reasons. The thing is that the possibility is there and when we talk about the data, and I alluded to that in the beginning, a lot of people do seem to not be aware that they're biased. But the important thing is that whatever you say on an explicit measure of the attitudes, or what you say when people ask you how much do you like different groups of people, your listener won't tell us anything about how you know about your spontaneous reactions. Those are two entirely separate questions, right?

I can ask you how do you feel about Black people, white people, Americans, Canadians, Europeans, people from China, and whatever you say there won't tell us anything about whether you know what your first reaction is, because you may reject it, you may not reject it, you may not be aware of it. We don't know any of that when we ask you that. So, the thing we have to do if we want to know do you think you have negative associations with any of these groups is we have to ask you that. And that sort leads to the research that I did is as simple as that. We just gave people the stimuli on an IAT and asked them, "If we measured your reaction, is it gonna show?" And it's quite easy to observe your own reactions and that very much speaks against the idea that that is something inaccessible happening there, going back to that definition of unconsciousness, so when you look at the stimuli most people can tell what their reaction is going to be and they can say that, and they still don't base their judgment about the entire group on that first reaction.

Andy Luttrell:
Yeah, so yeah, that's exactly where I was hoping to head, which is that so far we've kind of just been saying hypothetically we don't have to buy that these implicit reactions are unconscious, right? But I don't know, maybe they are. And so, the work that you've been doing is to sort of like really get at that question of is it possible that people are aware of these things, that the definitions we've been throwing around would say, "No, there's no way that people would be aware of this." And so, you hinted at it here, but what do some of these actual studies look like and what do the results show that we can say like, "Yeah. Okay, we can be skeptical that people are absolutely unaware of these things."

Adam Hahn:
Right, so what we do is we show people the stimuli they're gonna see on the IAT and give them the category label that they get towards them, so in the most famous example, that's why I keep getting back to it, not that there's not a lot of other stuff you can apply this to, but for Black and white people you show people the stimuli of Black and white people you use on the IAT and you simply ask them what is your reaction towards these pictures. And if we ran a test, that was our first study, what would it show? Would it show a more positive reaction towards the pictures on the right or the pictures on the left? You tell us. And then we let people predict their scores on five different IATs for all sorts of different social groups.

And what we found in those first studies is at least that people can predict the patterns of their supports very accurately. So, there are a lot of qualifications to those results and that takes a lot of time to unpack every part of it, but the sort of first line simple result is people predict the patterns of their scores very accurately. They will be able to tell you, "My reaction towards these people is more negative than my reaction towards these people, and these people, and these people," so

across sort of eight different groups or five different IATs that we use, people can tell you their relative results.

Andy Luttrell:
And which is just like if I directly ask you, you can still actually on this sort of explicit thing say something that actually is very consistent with the score you're about to get on my implicit bias game.

Adam Hahn:
Exactly.

Andy Luttrell:
So, I know one of the things that matters too is that part of the challenge is maybe that we often talk about bias in a very abstract way, and sort of like, "Do you have a prejudice generally about this group?" As opposed to the IAT, which is very specifically like you see a person's face and you have a response to it. So, in what way does that affect how we think about these processes as inside or outside of awareness?

Adam Hahn:
I'm not sure I understand the question.

Andy Luttrell:
So, I guess I thought you had some data on this distinction between if I ask you generally about… like people might lack some awareness about these sort of more abstract attitudes, whereas they can more… What you just said is when I show you a picture of a person, is that right? And what would your reaction to this person be? As opposed to like if I showed you… I won't show you yet, but if I show you generally pictures of people from this social group, what reaction do you think you would have? Does that matter or does it not matter?

Adam Hahn:
So, it matters to show them the actual pictures, which is sort of an interesting effect that we're still trying to explain. So, if I tell you we're gonna run a test on you where pictures will show up on screen and it's gonna measure your bias about these different groups, now tell us how you're gonna score, and we don't show them the actual pictures, they're not as accurate as when they see the pictures. And then we also try to switch out those pictures and show them any pictures, just not the ones that are gonna be on the test, and they still get better than without pictures. And so, sort of my current interpretation or intuitive interpretation of that is when you look at pictures, you can feel an affective reaction, but we seem to not be good at predicting that, so there's a lot of literature on what we call affective forecasting. We are generally bad at knowing how we are going to feel in the future in reaction to something.

Andy Luttrell:
Actually, it's sort of a compelling checkmark on the side of these things are unconscious in the sense that I think that's often what we mean, right? We say that we have these sort of global reactions in favor of some groups and against other groups, and we don't realize that we have those, and so some critics might say like, "Okay, yeah. Sure. If you put me directly in the face of

some specific person, it's hard for me to deny like actually how I feel in the moment. But if it's just like do you realize… Like the patterns of people's responses show that there are these abstract social categories that people have responses to and that's usually what we mean when we say you have an implicit bias, or you have a prejudice. And that seems to be the kind of thing people are just less in tune with, right? Is that kind of what these data are showing?

Adam Hahn:
Yeah, so we sort of seem to be unaware that we react negatively in a certain way, and part of it might also be that in the real world, we have a lot more context which comes into sort of this bigger question how valid… I mentioned earlier that the IAT is the most reliable, but there's a lot of debate on what it is picking up, and a lot of the IAT situation is of course artificial. You're only seeing pictures and the only information you're being given about these pictures is they belong to this social category. These are Black faces, and these are white faces. In the real world, you see people with context. When you meet people in a certain context, these are your colleagues. These are friends are friends. This is a stranger on the street. And all of this will factor into your reaction, and you may not primarily just react to the race of the person. That might be a result of the artificial setup of these studies.

So, that might be one of those things. But generally, what the IAT is picking up is just your reaction towards faces, and if we set up the explicit question to measure that, if we ask you, "Here are those faces and all we want you to know is how do you react towards these faces which we have now labeled as Black, and these faces which we have now labeled as white," then you will be able to know what your reaction is. In the real world, that is much more complicated, which may also explain why people are "unaware" of these biases before they come to these studies.

Andy Luttrell:
So, when you have people predict their reaction to a specific person or report their affective gut reaction, you're labeling race? That's a prominent-

Adam Hahn:
We tell them what the IAT… We ask them exactly that question that the IAT is asking. So, these are pictures of Latinos. These are pictures of Asian people. These are pictures of children or adults. What is your reaction towards these pictures?

Andy Luttrell:
Yeah, so it's very clear what you're trying to gauge their reaction to. So, at some level it does capture some intuition about these category-level things, but probably they're just responding to the salient features of like, "Oh, this person in front of me, do they make me feel comfortable, safe, open, happy, or not?"

Adam Hahn:
We give them a whole bunch of pictures. It's usually 10. It's not just one person. I think with one person the reaction might also be different. And one thing we're beginning to find out, that the data are sort of inconsistent so far, is that the contrast in the IAT seems to be a big factor of it. So, when we have people predict their reaction just by first looking at pictures of Latinos and Black people and all of those groups in isolation, even 10 at once, they become less accurate. So, what

the IAT is picking up is sort of a contrastive reaction of one category versus the other, and you have to ask it in the same way. So, part of that is something we know in methodology if you sort of match the methodological features of two types of measurements, they will be close, more closely related, but we try to set it up in what we call a two-by-two design, cross design, so this is a study one of my grad students, Alexandra Goedderz, is running, where we have people predict their reaction either by looking one group at a time, as I said, or by groups in contrast, and then they also complete IATs in contrast or so called single target IATs, where you're just supposed to categorize pictures of white people or just pictures of Black people with good or bad.

And what we find is two main effects. So, it's easier to predict the contrastive reaction, but it's also easier to predict a single target IAT with a contrastive prediction, so contrast always creates a stronger reaction that's easier to read if that makes sense.

Andy Luttrell:
Meaning people are pretty… Like biases come out very strongly in all of these domains when people are pitting one person against another, right? So, when it's a preference more than just like my reaction to you, right?

Adam Hahn:
Yeah.

Andy Luttrell:
Yeah.

Adam Hahn:
Well, and 10 people versus… One group versus another group. Not necessarily one person versus another person.

Andy Luttrell:
But at any one point, it's one person versus another person, right?

Adam Hahn:
No, it's always all 10.

Andy Luttrell:
Oh, in an array. All of these people compared-

Adam Hahn:
Yeah.

Andy Luttrell:
Oh, interesting. Okay.

Adam Hahn:
So, you really perceive them as a group.

Andy Luttrell:
Is this true in the original? Because what I was wondering was like when you gauge the accuracy of people's predictions, it's not my reaction to you, like my gut, what I say my gut reaction to you is compared to in this IAT what my reaction to that face was. But it's sort of just like across faces from one category, my prediction of what my reaction will be, that sort of average predicted reaction corresponds with the average bias I show on the IAT. Is that right?

Adam Hahn:
Well, we only ask people to do one prediction. We never thought of sort of making that more reliable by asking them multiple times, so we show them all pictures at the same time, so 10 pictures-

Andy Luttrell:
So, that's always been how it works.

Adam Hahn:
Yeah. And we just get the one prediction and that predicts your IAT. They might look more accurate if we ask them several questions. It's sort of we didn't know how to ask the same question twice without it seeming weird to participants. Is your reaction towards these faces more positive than these faces? That's kind of it.

Andy Luttrell:
Yeah. And that detail helps speak to the categorical versus specific nature, because what I was imagining was as you were describing it, right? One person at a time I go like, "How do I feel about you? How do I feel about you?" And sort of do those biases, like again, that feels like something people could very easily be in touch with their initial reactions about. But it is interesting to say like, "Oh, no. Show me all 10 of these faces together and I can report that this collection of faces does not trigger in me as positive a response as this collection of other faces." And so, that's interesting. One question, right? When you see these people versus these people, which one do you favor? Which group do you favor? And that is predictive of what they're doing on these sort of very quick judgment activities.

Adam Hahn:
Yeah. And again, and we tried to do it one at a time, and I should be cautious because those data are unpublished and we're finding some inconsistent stuff, which we will all publish in one, but when we do ask them for one group at a time, they're not as accurate. They're not completely inaccurate, but they don't feel it as strongly, which is another thing that might explain why people are so unaware of their biases, because we meet people one at a time on the street, and it's only in contrast that people notice that their reaction is more positive towards some groups than other groups.

Andy Luttrell:
Yeah. That's interesting. So, another piece of this puzzle that I wanted to bring to you is… So, I said one reason why people have often said that these biases are unconscious is just because these implicit and explicit ones differ, but another sort of more anecdotally is that if you've ever taken a class having anything to do with implicit bias, you've taken this implicit association test on the

website which at the end gives you some information about yourself. And anecdotally, people report being like surprised to be like, "I had no idea. You're telling me I have a preference between those?" Which, if that's happening, sort of suggests again, like writ large, people go, "Yeah, these must live outside of awareness," because people wouldn't be surprised to hear something they already knew about themselves.

So, either they're lying when they say, "I was so surprised," when inside they're like, "Oh my God, they caught me," or something else is going on. So, how can we make sense of the surprise people feel if these things are not absolutely outside of our awareness?

Adam Hahn:
So, you're nicely leading to a paper that we hope is in the last round of revisions. This work is surprisingly hard to publish. But we actually tried to ask people after IATs how surprised are you at your result and we tried to show that the reason is everything I've been talking about in this podcast. So, the fact that you don't pay attention, and so we do what we've been talking about. We show people these pictures all in one slide and say, "Just think about how you're gonna score on this IAT before you take it." And another group doesn't do that, and lo' and behold, we always get very significantly lower surprise reactions when we make people reflect on their biases first.

And so, of course one of those explanations could be that people just pretend to be surprised when they haven't paid attention and then the explanation for the attention manipulation, lowering that surprise is that we're sort of changing what they think is the desirable answer. So, their desirable answer is to pretend, "Oh, I'm so surprised to be biased." But then when we asked you ahead of time, you're like, "Oh, now I don't have to pretend anymore." But we try to rule out that explanation by doing a condition without attention where we also asked them, so ahead of time we were like, "Don't you think you're gonna show bias?" But we don't show them any pictures. It's just like this abstract racial bias, as we said before, without pictures, without a chance to observe your own reactions, and that doesn't have that effect. So, it's not about making people admit to their bias ahead of time. It's really seeing the pictures. And then we tried other conditions where we showed them the pictures and tell them to pay attention, but we don't make them make a prediction or anything, so we don't make them admit to bias. We just tell them, "You know, you're gonna do a test on these pictures. Maybe you want to think about how you're gonna score, but you don't have to tell us."

And even then, they're less surprised, so it's really about the attention to your reaction towards these specific stimuli that you somehow don't do in everyday life, so people who don't do that are surprised at their IAT results. So, again, going back to your question, people who say this is evidence of unconsciousness, people who are so surprised when they do IATs, they're right. Most of us are unaware of being biased and that's why we're so surprised, so there is… Again, the problem is that when you say the IAT measures unconscious bias, it sounds like it measures inaccessible bias that you cannot know, and that's just not true. You just haven't paid attention.

But the interesting part is that most of us haven't paid attention throughout all of our lives, which is why racism and inequality, and all these things are so pervasive, because most of us don't seem to be paying attention at our minor little biased reactions all day, and that's why the IAT is so

surprising. But not because it measures something inaccessible, but because it measures something you likely haven't paid attention to.

Andy Luttrell:
In those studies, did you always give them some feedback about which they could be surprised? Because I could see just the process of taking the test, you go like, "Oh." That's when you notice, right? The feedback is just sort of… I guess the question is what is surprising to people, the actual feedback they got where they go, "I didn't realize I showed a bias on this test and you're telling me I did," or just sort of the global experience of realizing, "Oh, I do seem to have shown some preference between these groups and I hadn't thought before about how I might have done that?"

Adam Hahn:
I don't know which one of the two it is, at least I don't have any clear-cut empirical evidence, but anecdotal evidence suggests that people still don't… They're still surprised. Noticing yourself having trouble on the IAT somehow doesn't change anything because people don't… They often don't attribute those challenging parts of the IAT to the race of the stimuli. So, when it's a race IAT and you ask people afterwards, they will tell you, "Yes, it was harder to associate this group with good than this other group," but they will claim it's because of the order setup of the IAT. That's by far the most common answer we get. They always say, "Oh, it's because you made me associate white people with good first and then I had a hard time switching."

And they're right. This effect is there, right? There's an order effect. It's just tiny. They would have still shown a bias if you had done it the other way around. And so, that's not the case, so that's also another explanation of why so many of us seem to be unaware of our biases, that we often attribute reactions that we have to other things than the race of the person. And so, that makes for a very comfortable way of going through life to be like, "I just don't like some people and I like other people. It has nothing to do with their background." And people do that apparently even while they complete IATs, right? They see how challenging it is and they're like, "Oh my God, this horrible setup of this test. You made it so difficult. It had nothing to do with the background of the faces." And we know empirically that although it's true that the order plays a role, it's not nearly as big as to explain these biases.

Andy Luttrell:
Yeah. It's sort of like you could take that as evidence, like, "Oh, people don't realize they have these biases," or they've deliberately chosen not to believe that those biases are legitimate. That's different than them not knowing. It's more of a, "No, no. I've made a choice about what to think of this."

Adam Hahn:
Which makes it so interesting that all of this goes away if you just show them pictures and encourage them, think about it. Are you gonna be biased if you do this test? And suddenly all these order explanations are gone, right? There's no… Once they realize, "Oh, my reaction is more positive towards one of these groups than the others," they totally move away from, "Oh, it was just the order of the test," or anything.

So, there's also a large part I guess within our subset doesn't want to be biased, which I think is a good thing, like we always sort of portray these social desirability biases as something negative or dishonesty, but I think there's… For a functioning world, we want people to act in line with their values, and their beliefs, and not their spontaneous first reactions. Like if you break it down, that would mean people don't go to the bathroom when they need to pee. We want people to hold back their impulses and sort of live them in a controllable manner, right? That's how a society functions. And in the same way, your spontaneous reaction towards categories shouldn't be the thing that guides your behavior. It should be your controlled and thought through reaction. So, it makes sense that people sort of don't want to pay attention to every little impulse that happens within themselves and rather listen to their values.

Andy Luttrell:
I just wanted to clarify something about the surprise finding, which is like the simplest way you could erase the surprise would be to tell people in advance, like, "You're gonna take this test and it's gonna show you you have a pro-white, anti-Black bias." Then people take it, they get that feedback, of course they wouldn't be surprised because you already told me that's my result. But really, what you're doing is sort of saying, "I'm not telling you you're gonna get that result. But just by me sort of hinting at here's what you're about to do," people come to the conclusion themselves because they are able to access those feelings to say like, "Oh, I see what's about to happen." And then they do it and they get the feedback, and they go, "Yeah, that's not surprising anymore because I already kind of realized this is what was coming. Not because you told me, but because I'm aware enough to have guessed this would be where things were headed.

Adam Hahn:
Yeah. Exactly. And if you take it further, like at this point in time if you consider that zeitgeist, at least in America, I don't think anyone taking a racial bias test is not prepared to be told that they have a racial bias, so that manipulation is omnipresent already, so in our control conditions everyone is expecting to get biased feedback. And they're still surprised. And they're not at ceiling anymore, again attesting to this sort of spirit of the times, it's like I said. Everybody knows they must have some bias. So, that reaction on a seven-point scale is only at around four or 4.5 or five, so people aren't super surprised, but it goes down significantly if we first let them look at pictures. That's when they really seem to get, "I may have biases I have not paid attention to prior to this study."

Andy Luttrell:
It would be interesting to see surprise rates over time, like from day one of Project Implicit website to today, like if they had been asking people… I don't think they have. Is this surprising to you? Probably in 2005 or whatever it was, people would be like, "What? This is incredible!" Whereas now, they go, "Well, yeah. Everybody keeps telling me this is what I'm probably thinking." So, by way of wrapping up, I guess what I want to put to you is the question of because implicit attitudes, implicit bias, whatever, still continues to be called unconscious bias, right? It's hard to escape that. I guess really the question is do you feel like what you're finding is compelling enough that you really would suggest that we get away from this terminology of unconscious bias? Or do you go like, "Hey, train has left the station. What's the big deal?" Is it that wrong to call it unconscious?

Adam Hahn:
I wouldn't say the train has left the station. I still think there's so much we don't know about implicit bias, so I definitely don't think we've reached the end of history with explaining implicit bias to the public. We're very, very far away from that. And also, when I talk to social psychologists, especially those working within it, I think there's an exaggerated image of how much has reached the public. There's definitely very educated corners of society that don't know anything about this, so I don't think it has… It's very widespread but not in the sense that like every person and their grandmother has the concept of implicit bias. I don't think we're there yet.

So, I think there's a lot of room to explain things, so just to that what should we do, and is it terrible that people are calling it unconscious bias, well, like I said, there's a lot of it that is unconscious and there's even an aspect we haven't gotten to it yet, so most people, we call that social calibration, don't know how biased they are compared to others. And so, that's why we have them take five different IATs and predict their pattern because we're already suspecting that everybody thinks that everyone else is more biased. So, if you compare different predictions of different people to each other, they're super inaccurate because they're all convinced that they have the slightest bias possible, and the most biased person thinks it and the least biased person thinks it. Everyone who feels a bias then shows us the minimal point on the scale, and you can only see sort of the accuracy by having them predict several IATs, because they can predict their pattern, but they're not accurate in knowing how they are compared to others.

So, there's a lot of things people don't know. There's the lack of attention, the fact that most of us seem to be walking through life without paying attention that we're biased. There's the fact that we attribute our biased reactions to 100 things, but never to race or those things that are problematic. There's the fact that these biases happen in an IAT, happen out of context. In a real situation there's a lot of context, a lot of other things that also trigger a reaction that might sort of drown out the reaction to just the race of a person and all of these things, so there's so many things that people aren't aware of and that make the real world a very different place where you may not be able to observe your biases in the same way that an IAT would observe them.

So, all of those things would suggest that it's okay to say it's measuring unconscious bias, because it's measuring a bias that you very likely don't know. The reason why I'm still fighting against it, or I think it's wrong is because I still think that people hear inaccessible when you tell them you're unconsciously biased. And there's some very intuitive research, Cameron and Payne showed that people think you're not responsible for an unconscious bias because how would you? You're not even aware of it. So, if someone makes biased decisions and we tell them that can be predicted by unconscious bias, and people say, "Oh, well there's nothing you can do about it. No one's responsible." And I think that's a terrible conclusion to draw because A, as I've been saying throughout this whole podcast, it's very easy to observe your own biases and I think everybody should be held accountable to observing their own biases.

And so, that's sort of the problematic aspect to me with saying unconscious bias. It's not entirely wrong depending on how you define unconsciousness, but what people seem to hear is we have inaccessible biases that we don't know and that we can't know, and that is just not true. That is just not true. We are all able to observe our biases all the time and it's hard for me to imagine a world where we say we have unconscious biases and still everyone is held accountable for them,

and that's the problem with saying unconscious bias to me, so that the lack of accountability that seems to come from it and… Well, and this idea that we're not encouraging people to be enlightened and observe their own biases, but that we're telling them to take a psychological test, which is also… The IAT is a noisy measure. It's not a perfect measure of your spontaneous reaction. It's at best an approximation of them.

And it's fun for researchers to have that tool. I've been finding a lot of interesting things. I wouldn't be able to say any of the things I'm saying without having used it, but I don't think this idea that the public needs to go to the Project Implicit webpage to find out if they're biased is true. Your best way to figure out if you're biased is observe yourself. And as long as we're talking about unconscious bias, people aren't… We're not encouraging people to do that. And I think that's a problem because that's where we, as a society, where we should want to be. We should want to accountably and reliably observe ourselves and check ourselves for our biased behaviors and that is not where we're going when we say we have unconscious biases. So, that's sort of my big issue with this.

So, yeah, it's not entirely wrong, but it has a lot of implications that are at least ethically and morally questionable and that might not be what we want to achieve by doing this research, and so from that perspective I think it's important to get away from the presentation of unconscious bias.

Andy Luttrell:
That's great. So, putting you on the spot a little bit. If someone were to say, "Okay. Well, if implicit bias isn't unconscious bias, how do we define it?" This is a giant question and I understand that I'm forcing you to take a stand on a very contentious part of a research field, but if you were to say, "I'm gonna write the textbooks. I'm gonna tell you on the left-hand column, in bold, implicit bias: …" How do we define it? What is it that we're talking about?
Adam Hahn:
So, I've been, and you could call this the easy way out. I've been using a very methodological definition. And so, I say implicit evaluations are evaluations that I see on a computerized reaction time measure and explicit evaluations are evaluations I see on a measure that you have full control over. And it's actually a surprisingly large amount of people in the field use it that way. It's usually the people who don't do research with implicit scores that don't use that definition and then criticize, "Oh, you don't even know what implicit bias is." I'm very open about I don't know what exactly it's representing. We're still finding out. I think it's interesting to find out what it's measuring because we can replicate these results. It's a reliable measure and so on and so forth. But I don't know what it is. I can't tell you.

We have these working definitions of it being a spontaneous reaction, but that would only apply to prejudice reactions. So, we also have stereotyping IATs, right, where we measure whether people think that women are associated more with family and men with career, and we just are also in the last stages of publishing a paper here with another grad student, Zahra Rahmani Azad, and she found that, and that cannot be an affective reaction, so even if you wanted to define implicit bias works on picking up spontaneous affective reaction, that wouldn't be true, because they're also out picking up spontaneous semantic or sort of belief reactions.

And so, in that sense it's just really hard to say. I just think the whole sort of bashing of if you don't know what you're measuring you shouldn't be researching it, I think that's a very anti-science attitude, so I don't… I think as a humble, modest scientist, you should be researching things you don't understand yet. That's the point of doing research. So, this whole I should have defined what implicit bias measures measure ahead of time, and since I don't I shouldn't be doing research, that really upsets me when people say that. I think we should be doing research on things we don't understand, and I fully admit that we don't or at least I don't fully understand what implicit bias scores are picking up, but they're picking up something very different that explicit bias scores, and explicit evaluation measures, and explicit stereotyping measures, and we do know some things about what those things are. They're most likely some spontaneously activated mental content. That's sort of one definition I can agree to. It measures spontaneously activated mental content. That's what I tell my students and what I write in my textbook chapters when I write for textbooks.

Andy Luttrell:
And we have no special reason to say it's unconscious. Sure, it might be, but it would be wrong to predefine all of the reactions we get from these tests as inaccessible reactions.

Adam Hahn:
That would definitely be wrong, and they might be unconscious for some people some of the time, right? You would call that preconscious. The point of the problem with defining it that way is that preconscious is by definition just a state, so it might apply at one point and not another point, so you can't say that we're measuring preconscious attitudes because we might or might not. They're conscious to some people and unconscious to others.

Andy Luttrell:
Well, this has been super helpful and very cool to hear about, so I just wanted to say thanks for taking the time to walk us through this.

Adam Hahn:
Thank you for having me.

Andy Luttrell:
Alright that'll do it for another episode of Opinion Science. Thank you so much to Adam Hahn for talking with me about his work and the bigger picture. If you want to know more about Adam's work, I'll leave some links in this episode's shownotes to point you in his direction.

If you're not already subscribed to Opinion Science, look into that and make it happen. OpinioSciencePodcast.com is the place to go for transcripts, past episodes, links to things, etc. And if you're feeling kind, which you are, right?... Leave a nice review of Opinion Science at Apple Podcasts or whatever app you use that also lets you rate and review a podcast.

Oh, also, final note. I often make a point of forewarning you when I'm under the weather to explain why my voice isn't all the way there, but I have a one-year-old in daycare who brings home germs constantly, so these episodes where I'm stuffed up or something are basically all the time now. Like this one. And you know, I don't even know if you can tell. But this is your final notice.

Okeedoke, that's all I've got for you. Come back in a couple weeks, and I'll have more Opinion Science for you. Buh bye…