Robots and Racism

[00:00:00] **Christoph:** I'm living in Christchurch, New Zealand. On March 15th, we experienced the worst terrorist attack in recent New Zealand history. We lost 50 people to racist violence. I'm struggling with coming to terms with the attack. Racism is a major problem, and I don't have all the answers. Last year, we published an article on robots and racism compared to the damage that racism does amongst most humans.

[00:00:27] Our study seems trivial. Still, if anything at all, I hope that our study brought the topic of racism to the attention of the human robot interaction community. We were certainly not the first to discuss racism and even sexism in HRI. Today we're going to talk about these difficult topics.

[00:00:49] This is the Human-Robot Interaction podcast.

[00:00:55] I am your host,

[00:00:57] Christoph Bartneck.

[00:01:05] Racism is a very broad term and thus requires some explanation.

[00:01:09] **Kumar:** The The term racism is a bit complex, partly because there are multiple meanings to it. My name is Kumar Yogeeswaran. I'm a social psychologist at the University of Canterbury in New Zealand. Typically when people use the term racism, they're thinking of showing favoritism towards their, say one group over another racial group.

[00:01:29] But I think the terms in the academic literature that I think are more appropriate, often start to distinguish the ways in which people can show bias or prejudice towards one group over another. So to give you an example one could think of dislike towards one group being at a very subtle level. So people say have implicit bias in favor of one group or against another group while it could also be much more blatant where a person knowingly say discriminates

against another group and the term racism kind of combines all of these things together.

[00:02:06] So sometimes we were talking about behavioral tendencies where someone's actively discriminating against one group. Sometimes we're just talking about people's. Say preferential treatment for someone in one group over another other times we're talking about stereotypes or beliefs about a group and all of this just gets combined into the same term racism.

[00:02:25] So instead of the literature, we often make a distinction, both in terms of, are we talking prejudice, which often has more to do with dislike, antipathy, hostility, and discrimination, which is usually behavioral. And within that, we separated into the more, very blatant forms, which would be extreme hate the most extreme end of course, you would get genocide itself, intergroup conflict at that level as well.

[00:02:52] Also a form, but then usually what we encounter on a more day-to-day basis, especially in Western nations, tend to be more everyday casual aspects of it. So sometimes it can be direct where someone says, I don't like to hire people from a certain group. It's illegal, but that would be an example of kind of blatant discrimination, but then you can also get to more subtle forms. So for example, aversive racism as a framework within psychology, that's used to explain how people can sometimes engage in discrimination specifically in contexts, where it's easy to rationalize. So these are people who are very well-intentioned.

[00:03:34] In contexts where it's almost ambiguous and it's easy to rationalize their behavior, they're more likely to discriminate only in those contexts. So say if you have a very weakly qualified job candidate who's African-American and white American at the same time that they're considering they equally show a lack of interest in both of them. If the candidates very strongly qualified and they're African-American and European American at the same time, you see again, they don't show a difference, but if the qualifications are ambiguous, then they tend to prefer the person who's White American or European American in that context.

[00:04:07] So it's easier to rationalize with ambiguous qualifications. So aversive racism became one of the early indicators of what was a relatively subtle form in which emerges. But since then, the fields kind of moved even

further with developing better and better tools which involves the reaction time tools or the neuroscience measures that allow us to better understand.

[00:04:31] Kind of very subtle biases that people may have. So it becomes a lot easier to use those tools to be able to capture what are almost automatic tendencies that people can engage in. So the short answer is that racism as a term is a bit problematic to define because it can almost be over encompassing. People are usually implying many different things with it.

[00:04:54] And the term can almost be overused in some contexts where almost anything can be considered racist. So problem with that becomes that. Now you have even a white nationalist groups that don't like, say that they're not racist, even though they express views that indicate say white supremacy, they just see it as "Well, I just love my group. That's not racist". And I think that term can almost be loaded to the point where nobody wants to be associated with it. So instead it helps to kind of make these finer gradations to say that, well, there are different ways in which these biases can manifest themselves rather than just throwing a general term.

[00:05:32] **Christoph:** All right, in 2012 ,for the Friederike Eyssel and her team experimented in Germany with robots being either a member of their national group or being from a different group.

[00:05:42] **Friederike:** We have done several experiments where we manipulated the social category of a robot, meaning the social group that. robot allegedly belong to. We manipulated the name of the robot by telling participants this would be a German robot called Amin, typically German name, and this robot would be fabricated in Germany.

[00:06:07] Or the other group of participants learned that the robot would be called Amman a typically Turkish name. And this robot was built in Istanbul at Istanbul University. And when people learn this very subtle information, because they were just confronted with it when looking at some pictures of this allegedly newly developed robot, they did something that people tend to do when they judge Other groups. We looked at German participants and these German participants evaluated the Turkish Robot prototype. And we were wondering whether the mere categorization of a product. A robot as German or Turkish would lead to an effect that we call in-group bias. That means that people would evaluate the robot that belonged to their in-group more favorably than the one that allegedly belonged to a social outgroup. So

we assumed that the German robot will be rated as more mindful, as more capable as more socially- able and even better in terms of its design compared to the Turkish prototype. And in fact, that is what we found to our surprise at German participants, derogated the out-group product and rated the German version of the same product. They saw the very same picture of the very same robot that just was given a different name in a more positive way when it was built by an in-group member, so to speak and had a German sounding name. We interpreted that. yeah, as a replication of what we commonly find within the human-human inter group context, where we also rate our in-group more favorably than the out-group and discriminate against the out-group by assigning traits to them that are less favorable and reserve all the good things for the in-group.

[00:08:18] **Christoph:** The next experiment of a team took place in the US.

[00:08:22] **Friederike:** Having looked at such an experiment. We were just following this line of research and extended it by having in a second experiment, white American participants rate a robot. This time we didn't play around with the name of the robot or the location where it was produced, but we rather changed the appearance of the robot by having one robot Look a bit more, "Caucasian" and the other robot, having a bit of more darker skin to represent the social out-group in the US American context, maybe an African American instanciation of a robot. So by coming up with such an intergroup context, we were wondering whether we could replicate the effects in the German Turkish version.

[00:09:14] Now looking at white Americans and their reactions to allegedly Caucasian versus out-group robot prototypes. And this time we asked our participants about the mind perception, the attribution of agency or experience to either one of these Robot prototypes. And this time again, we were highly surprised because the white American participants reacted differently from what we had expected.

[00:09:44] These, participants actually attributed equal levels of mind, agency or experience. So the robots capability to form plans to react emotionally, to display emotions. They rated these robots likewise and it didn't differentiate or discriminate against the out-group version. However, we had measured participant's level of modern racism, so their anti African-American beliefs. And when you then differentiated participants, in terms of people with high versus low levels of prejudice, against African-Americans we found exactly

what we had predicted. People with high levels of racism rated the out-group robot differently from people with low levels of racism against the social group.

[00:10:44] And so the attitudes they brought into the lab had a significant impact on their evaluations of the robot. And we were actually pretty happy that we had this extra covariate in, in the study, something that you commonly just measure when you're a social psychologist by training, but not a roboticist in a sense.

[00:11:08] And it helps us to understand why people would differentially evaluate these two robot prototypes while when you would neglect such attitudes that people just bring with them, we would not have found the effect after all, you know, we would have thought, oh no, that's not happening. We don't face substantial societal problem.

[00:11:33] Maybe because people also use social categories that they are dealing with in human- human interrelations when they think of social robots.

[00:11:44] **Christoph:** There's not only a racial bias, but also agenda bias.

[00:11:48] **Friederike:** So gender, in terms of the data that we have on gender, we have found for instance, that indeed, certain robot features like hair length or body shape can activate a gender stereotypical knowledge structures. And we would, in that sense also use stereotypes when judging male versus female robots. So attributing Gender stereotypical traits to a male versus a female robot in terms of agency, for instance, that is reserved for more male robot, West sociality or warmth, it's reserved for the female type. The suitability for typically male or female jobs is also influenced by the gender of a robot. We have further research conducted by Natalia Reich-Stiebert and myself, where we see that in the learning context, this can have an effect when your robotic tutor is male versus female, and this interacts with the type of learning task you do, whether it's a gender, typically male task you do something like the robot teaches you binomial formula on mathematics compared to a robot that teaches you verbal skills and stylistic means, forensic rhetorical means when taking that into account, you sometimes see that for instance, a mismatch between robot, gender, and gender, typicality of a task can lead to more positive outcomes in terms of motivation to learn with the robot.

[00:13:25] So gender plays a role, not, not only in terms of what the robot looks like when you take into account the physical appearance, but it also matters in terms of what job or task the robot is doing. So we have to consider that I would say in terms of actual behavior towards robots, because you were mentioning Megan Strait's research on aggressive behavior towards female robot prototypes, we have so far not done research in similar lines of research. We have all we need and that's, I think what makes Megan's research very attractive. We need actual research that looks at behavior direct behavior towards the robot and not just attributions of traits. Or perceived suitability for certain tasks. So the actual interaction with the robot is key.

[00:14:17] I would just say that much more research has to be done in that regard. I think in the realm of gender, we also would need to take into account sexist attitudes for instance, or attitudes towards sexual objectification. There are scales out that measure the degree to which males and females tend to objectify other humans.

[00:14:40] So we know for instance, that people who tend to have a high proclivity to objectify others, that means to instrumentalize them, and to treat them as objects as mere body parts for physical pleasure. For instance, people who do that to other humans also tend to be more interested in owning sex robots For instance. That's data that my students, Julian Anslinger has gathered. I also believe that when I hear that robots are targets of actual aggression and particularly female robots, I would believe if you investigate further the effect of sexist attitudes or proclivity to sexually objectify humans, then this could clarify the whole picture even more and give us insights into the psychological mechanisms, underlying such behavior towards robot because I think that is what needs to be studied further. Why are we actually doing this?

[00:15:45] **Christoph:** Megan Strait, University of Texas Rio Grande Valley also studied how people reacted to human-like robots. Their starting point was a study on the Uncanny Valley, but it took them on a path to investigate a racial and gender biases. They analyzed comments made to popular YouTube videos about robot.

[00:16:06] **Megan:** Yeah. So this is actually short reports bridging two papers. So the first paper was actually published in 2017 with the original idea was to look at the public responding to human-like robots with a specific focus at looking at how frequently we see the Uncanny Valley referenced in popular

discourse with the aim of trying to understand, to what extent it's pervasive in how people react to an emergent robots.

[00:16:34] But what we did was we, we identified a set, I think, of 24 robots where 12 of them were highly human-like, and 12 are more mechanic morphic. And we're trying to contrast people's free form commentary. And we went through trying to code for the theme of, to what extent we see the Uncanny Valley. But what we saw as we were doing that was that a lot of the commentary was in fact, very abusive and very objectifying.

[00:16:59] And so we revised our coding schema and ultimately that gave us the ability to see the frequency at which people, at least in public forums, specifically, in this case, YouTube exhibit more aggressive tendencies. So what we specifically coded was the frequency at which they invoked stereotypes, the frequency at which they utilize violent language.

[00:17:22] Threatening violence or encouraging violence towards robots and the frequency at which they objectify them typically in a sexual manner. And what we saw with that study is that similar to human social dynamics, there was a greater frequency at which people were dehumanizing. So encompassing all of those themes towards the highly human-like robots that were female gender.

[00:17:43] And then we followed that up with a second study to look at how this intersects with their racialization. And so this short Abstract that appeared in the HRI late-breaking reports was bridging this of focusing more specifically on greater dehumanization and thinking more critically about once we saw as to what we could do to analyze their differences in responding to the female gender robots that varied in their racialization.

[00:18:10] **Christoph:** What robots did you specifically look at?

[00:18:12] **Megan:** We looked at Bina48, Nadine and Yangyang who's racialized in our likeness of and a Chinese, I think semi-celebrity, women of modern celebrity. But we picked these specifically because they were apparently similar in terms of their perceived age. And they were also all female.

[00:18:30] **Christoph:** So how do people react to this different racialized robots?

[00:18:35] **Megan:** So, this is the core of our second analysis, which was published in the 2018 paper. Where, what we looked at similarly was the same themes. Then looked at the frequencies across the three racialization where Nadine who was racialized as white was subject to a significantly less dehumanizing commentary relative to both Bina48 and Yangyang, which again mirrors that what we see with human social dynamics. And so together, both of them seem to give at least preliminary evidence that there's this really automatic extension. So even though the data that we had collected was from people who knew what they were observing were robots there is seemingly right this automatic extension at the same bias where we see this gendered effect, where the, when we gender the robots As female they're subject to more dehumanization. And when we racialized that this compounds on the gendering to exacerbate the frequency.

[00:19:35] **Christoph:** Megan looked at the dehumanization of robots and people, but what exactly is dehumanization?

[00:19:41] **Megan:** Dehumanization broadly refers to ascribing less humanness to an entity. How we've used it in this context is that we have these very human like robots and the effect of people invoking stereotypes or threatening violence or objectifying them sexually is in effect dehumanizing of them. That is denying them less humanness than what their appearance prompts.

[00:20:07] **Christoph:** But a robot is not a human. So how can anybody de-humanize a robot?

[00:20:11] **Megan:** Yeah, so that logic makes sense if we didn't respond to robots socially, but especially in this case where the set of robots that people are responding to have appearances that evoke Attributions of humanness that simultaneously what we have is people perceiving entities, agentic entities that look human, that prompt attribution of a human trait, but then simultaneously engage in these behaviors that would be useless or meaningless if they weren't perceiving them as human.

[00:20:44] So those two things coupled together, the fact that their appearances are human-like and that people are engaging in these behaviors that carry no meaning if you're not perceiving humanist implies pretty strongly that it's possible to de-humanize entities that aren't human ontologically, but nevertheless, give rise to that perception.

[00:21:03] **Christoph:** Friederike was also curious about the dehumanization of robots and humans.

[00:21:09] **Friederike:** So I've become particularly interested in the topic of discrimination, against Other social groups, because before studying robots, I was actually quite interested in the question of why we dehumanize other members of social groups. And I've studied that for instance, in the German context with German and Turkish people, but also in the context of Germans and Gypsies and Roma people. And in that research that also has been published in social, psychological literature. We have found that there had been quite pronounced prejudice against. Roma people in the German context and from thinking about the psychological underpinnings of dehumanizing other humans, I have become more interested in why, however, we are interested in humanizing non-human entities.

[00:22:11] So why we strip off humanity of fellow humans, but at the same time, anthropomorphize non-human entities like robots. And so I started researching the psychological mechanisms underlying this and wondering whether the same motivations drive people. We ran a number of psychological experiments to understand better why people humanized robots.

[00:22:40] Why they attribute mind, agency, experience, emotion, typically human traits, human essence to non-human entities, and under what conditions, however, people would also deny other robot such essential humanity. So I try to bring up so-called into a group context within the study of social robotics. And since when I first started becoming interested in robots, I was very interested in the very simple manipulation of physical features of the robot.

[00:23:18] For instance, the study that I mentioned on robot gender, where we manipulated the robot appearance by just varying whether the robot had long versus short hair, that was a very trivial and very just a study that I did for fun, so to speak, but it inspired all this other research because I try to follow up on.

[00:23:41] On the very subtle things that you can change about a robot and where you could see effects that are aligned with what you commonly find in the human- human sphear. And because my interest was to essentially replicate these findings and to see to what extent the human social cognition would also match the cognition have it played when we deal with robotic social agents.

[00:24:10] **Christoph:** But back to Megan, what did the comments on YouTube include? Can you give some examples?

[00:24:16] **Megan:** I think we do include a few examples in the publications in general, a lot of the sexual objectification. Revolved around wanting to engage sexually with the agents and varied on the basis of their racialization, where similar to what we see with respect to Asian women, that there's a different way in which people sexualize Asian women and sexualized Yangyang than they did with the other two robot exemplars. in terms of the stereotypes, I prefer not to rehash them to Not reinforce them, but they are certainly things that I'd be willing to point folks to references if they wanted more information in terms of what they were grounded in. But a lot of what they were utilizing were common stereotypes that are grounded in historical and long-standing marginalization that makes them marginalizing by utilizing them.

[00:25:14] And then in terms of the physical violence, there's a lot of interest and perhaps one explanation of this is curiosity, but there's a lot of interest in physical destruction of the robots. And we specifically coded a separate theme to look at. Did the content seem more driven by fear of, for placement or fears about robot uprisings or along that lines To track to specifically separate out things that seem motivated in a distinct manner than from the dehumanizing commentary. So for those, it potentially is curiosity, but there was just a lot of interest in physical violence. So like punching the robots or there's a common meme about a will things blend.

[00:25:56] So there's that meme, but it's also very odd to invoke that meme when you're talking about a very human-like entity.

[00:26:03] **Christoph:** In 2017 and 18, we conducted two studies that investigated racism. in HRI. One of the methodological challenges have been how to measure a racial bias.

[00:26:15] **Kumar:** One of the major reasons is that when people began to look, so post-civil rights around the U S researchers were noticing, there is still plenty of contexts in which you would see unequal outcomes unequal levels of say, disparity of let's say, health outcomes, or housing, job hiring among other things, but they would often find that when you just ask people, you know, how would you treat people from these two different groups?

[00:26:41] They're like, no, of course I'll treat them fairly. And so it was almost like there was a gap between what people were saying and what we were seeing at larger societal level or even within a organizational level. So people became much more careful what they began to express outside. And this is where it became clear that if you were just going to stick with what was the traditional approach of just asking people direct questions, like how do you rate this particular group? We know that self-report is heavily prone to these social desirability biases that if you ask them, they think about, well, who's reading the. Do I want to sound like I'm bigoted and people then can alter their responses accordingly. So that was one of the major motivators for why people began to look at these more indirect approaches.

[00:27:29] The other reason for it seemed to have more to do with the fact that we began to realize in psychology, that people have this kind of system 1 system 2 processes that take place at the same time. So we can have these very controllable, deliberate thoughts and actions. Or we can have things that are much more automatic.

[00:27:48] So our cognitions and our attitudes can be much more automatic. We don't put much thought into it. We just engage in specific behaviors even, and making the, finding this distinction between automatic and control processes meant that the field began to say, well, we need to look at both. We can't just rely on asking people what they think, because there may be these other automatic processes that lead to different outcomes. So the tools were largely developed in response to both societal changes at one level, but also the importance of being able to understand these automatic processes alongside more control processes.

[00:28:25] **Christoph:** We decided to use the shooter bias paradigm for our experiment.

[00:28:29] **Kumar:** So the shooter bias was paradigm established by Josh Correll. And who's a PhD student, I believe at the time who was working with the supervisors. They were all quite horrified by the shooting of Amadou Diallo in New York city. So this is an African-American man immigrant of who was in New York city when the police were looking for a certain suspect to a crime.

[00:28:52] And when the police surrounded him, he tried to pull out his wallet to show ID, but they have mistakenly thought he was pulling out a gun and he was shot about 40 times. And this led people like Josh Corell to really wonder,

would it have been different if Diallo was white American male instead. So was it almost that his race influenced the way in which the police misperceived the object he was pulling from his pocket, the gun, as opposed to an innocuous objects.

[00:29:21] So that particular incident led to the development of the shooter bias paradigm. So Josh Correll and colleagues began to simply develop stimuli where they would have African-American men and white American men holding different objects in their hands. And the task of participants was to make these judgments of whether to shoot or not shoot in response to each of those stimuli.

[00:29:44] Since then there've been better ways in which people are doing this, of course, but the original idea really came from looking at the specific police shooting and other ones like it that were of quite a bit of concern, but Correll and colleagues really wondered if there almost this automatic bias that seems to come into play.

[00:30:02] And is it specific to say white Americans in particular, their data began to show that actually this was a tendency that both even African-American participants were showing the same kind of bias, which then suggested that there may be greater awareness of say cultural stereotypes that could drive the way in which each of us can show these automatic biases.

[00:30:23] So they developed this particular paradigm tested with different populations. Eventually they even. Police officers and SWAT team officers within their studies and found that interestingly, the SWAT team officers and police trained officers seemed to be slightly lower than the general population in these biases, even though we would think that they were actually worse, the university students actually work. But, but it did seem to be point that there was this automatic tendency for people to mistakenly. People were quicker to shoot African-American men with were armed compared to white American men who were armed and they were quicker to not shoot white American men compared to African-American men.

[00:31:04] **Christoph:** When I approached you with the idea for the study, did you think that it would work?

[00:31:09] **Kumar:** It was an interesting question because to me, the distinction between whether people carry this over like a human-level bias that people

have at an automatic level would transfer over to robots that people are looking at these robots and they know they're not living. They know they don't have race at a conscious level, but yet if they would show those biases, that would be fascinating because it would suggest that there's some sort of carry over tendency.

[00:31:37] Simply the color of objects even can potentially influence us. So I, I just found it a really intriguing question that was worth exploring.

[00:31:46] **Christoph:** In our experiment, participants showed the same racial bias found through its humans in robots. That was a clear shooter bias towards robots, racialized as black. did the results surprise you?

[00:31:59] **Kumar:** To some degree, I think the first study, I wasn't sure how much to place on it simply because my field's gone through a huge replication crisis. it's always good to consider each study as a stepping stone towards better understanding. But now that we've run a few studies and the results seem to be quite robust, it does really raise questions about I, I feel like more work needs to be done to look at kind of the generalizability of the effects to know that the degree to which say the anthropomorphism of the robot actually matters because the more human it looks is that the only context in which you see it, because then you could say the group, the human group stereotypes are being carried over much more easily, but it's, it's not something I fully expected. I expected a general tendency, but I think the results seem to suggest that this may be fairly robust.

[00:32:49] **Christoph:** In our second experiment, we added a brown robot to the setup and the shooter bias disappeared. What could this mean?

[00:32:58] **Kumar:** There's a few different possibilities. I think one possibility is that the, when you have a diversity of groups, people make social comparisons between those groups. So once you introduce a third group that people are creating associations about, it's possible that people's perceptions of say, The darkest colored robots now have shift as a function of what the, say brown-colored robots in this context, what associations were elicited in that context.

[00:33:30] So I think one possibility is that having those multiple groups leads to these kinds of social comparative processes that drive people's reactions. But the other possibility is that it has something to do with the brown colored robots would it work eliciting different kinds of stereotypes at the same time,

because the open-ended responses when people are asked to think, well, do you think this, what race do you think this robot is?

[00:33:54] They were thinking of multiple different ethnic groups at the same time, with very different stereotypes simultaneously. the stereotype of people that look like me that are brown. People of Indian origin tends not to be a violent stereotype, but people do have stereotypes of people, of Hispanic descent.

[00:34:13] If, for example, Mexican individuals that may be more similar to the stereotype about African-Americans. So it's possible that muddies up what brown is interpreted as which then has carry over effects for the entire task. But it, the other possibility of course, is that when you have a diverse number of targets that you get people to see variability in such a way that group-based associations become less relevant. So there's some work exposure to diversity that shows it has benefits. Social identity complexity, for example, has benefits as well at the individual level as well. So it's a bit hard to know. I think we need to do more work to really iron out what might be driving it.

[00:34:55] **Christoph:** In a future that features thousands of robots, could the diversity of the robots help us to overcome racism?

[00:35:03] **Friederike:** First of all, people have to become aware that they're actually making use of social categories and that they sometimes use stereotypes and these biases, their judgment and their behavior in terms of robots. The big question is if people are so prone to automatically use their stereotypes, How can we actually counteract this?

[00:35:27] Would robots be the force to do that? Maybe by deploying counter stereotypical robots in certain domains. Like in our study, we found that female robots are the best nurse bots. Maybe we should have male examples as well. Like we have male kindergartners or male nurses to kind of break these stereotypes and these ideas. Robots could be a means of actually at least working towards this.

[00:35:57] The whole history of research on stereotypes is more than six decades. We've been trying to really resolve that problem from a social-psychological point of view. And we have not gotten very far so I don't want to be too pessimistic. But even without the robots, you know, we've been trying to reduce bias between humans and social groups to reduce intergroup

conflict and implications that come along with our automatic ways of thinking. and automatic ways of stereotyping others. I think we need many more empirical studies to test and evaluate whether indeed employing counter stereotypical examples of robots could help in terms of human- human, psychological experiments. We have been trying to use counter examples to kind of. Do you stabilize these mindsets and come up with fresher ideas and to change stereotypes.

[00:37:01] And there are many paradigms that have been put out in social psychology to say to change stereotypes. We've also tried to use them in robotics, like imagining contact with a robot to change entire robot attitudes, like imagining contact with a human out-group changes stereotypes about human out-groups and associated prejudice.

[00:37:26] **Megan:** I haven't yet considered the idea of deploying robots for providing greater representation where we racialize them and then provide those platforms as a opportunity for exposure by people who don't have much diverse exposure to more diverse peoples. If we were to do that, there sociality would have to be much more advanced than where things currently stand, where the state of art currently sits.

[00:38:02] And especially one of the things that we see is when we deploy these platforms in the wild, that is like unsupervised that if we don't equip them with capacities to recognize these sorts of behaviors and respond, that's just making an easy target for somebody to engage inappropriately and to have no consequences to that, or have no oversight.

[00:38:25] And so that's more dangerous than asking, say people of color to interact with people who have very limited experiences and might not be the easiest folks to engage with. So I don't see that as something in the near future that is using robots to increase representation in terms of how they could affect society in a positive manner.

[00:38:49] I think there's actually a fair amount of work already. So for example, Malta Hyung, he has a lot of work at these really small facilitary effects. So having robots intervene in a conversational breakdown between two interlocutors, where if somebody hurls an insult, the robot jumps in to try and repair that context by calling out the person acting inappropriately.

[00:39:14] So that is one demonstration that he's done some testing of, of how should the robot enter on a fight, but was actually just that HRI presenting several studies that she's done looking at similar usages, where one example was she had a garbage bin robot where it was designed to try and encourage people to pick up litter or pick up after themselves and throw it away where it utilized this proactive behavior.

[00:39:41] And that seemed to be influential at least in public settings for getting people, to getting people to actually put trash in the waste receptacle. So I think those are all very interesting, even if these single demonstrations and that they're all these minimally anthropomorphic robots, but that are able to promote some behavioral change in certain ways.

[00:40:04] That's potentially powerful, potentially really powerful if we extend it to the realm of social behaviors, such as how people are treating other people.

[00:40:14] **Kumar:** Possibly, the literature on diversity is very mixed and messy. Generally that literature tends to show that often it's not just about representation that has benefits for sure at practical level. But, so if you have a diverse workforce, for example, whether that actually translates to better outcomes is something that's a bit contested within the literature.

[00:40:39] It's not a clear benefit. You can say the benefits are more indirect, but I would imagine that simply having that exposure to various say robots that look like people from different groups tends to paint a picture of our society being much more complex that it built, Our society doesn't belong to any one single group.

[00:41:02] It's a pluralism and the face of the robots essentially map on to what actual human society looks like. So if I had to guess, I would say it may have the benefit of simply being, looking like one of us, as long as those robots are not developed in the sense that we reinforce the stereotypes by creating robots that look like one group that serve, say cleaning sort of tasks and another group that seems to be the robots that have the better kinds of roles within our home environments that only can reinforce the associations people have with certain groups and specific traits. But if we do just think of it as robots that are complex and they represent different kinds of roles it might Break some of these automatic associations we have if sufficiently look like members of certain group.

[00:41:50] **Christoph:** Racism is a problem that exceeds human robot interaction research. But even if there's only a small chance that we can make a difference, that we can reduce racial bias in our society. I think we do have the obligation to try. Thank you for listening to the Human-Robot Interaction podcast.