MIT CSAIL Alliances | Ruth Rosenholtz CSAIL Alliances Podcast 3

Welcome to MIT'S *Computer Science and Artificial Intelligence Labs Alliances* podcast series. My name is Steve Lewis. I'm the assistant director of Global Strategic Alliances for CSAIL at MIT. In this podcast series, I will interview principal researchers at CSAIL to discover what they're working on and how it will impact society.

Ruth Rosenholtz is a principal research scientist in the Department of Brain and Cognitive Science at MIT CSAIL. She received her PhD in Electrical Engineering and Computer Science from the University of California at Berkeley. Ruth's research focuses on visual encoding, particularly in peripheral vision, and its implications for visual performance in theories of attention.

Her lab studies a wide variety of visual phenomenon, taking a three-pronged approach, which includes computational modeling, behavioral experiments, and applying models an understanding of human vision to applications such as image compression, design of user interfaces, and design of information visualizations.

Ruth, thanks for your time today. Well, throughout the pandemic video conferencing platforms have become a part of daily life for so many people. As a vision scientist, what is intriguing about video conferencing and what makes you want to research it?

Well, so video conferencing has really enabled so much stuff during the pandemic. We could have remote work meetings and remote education, doctors visits that were remote, and even social events. But as a vision scientist, and particularly someone who studies vision for applications, I think that we could probably improve the experience through better design of the systems and better technology.

Many people have been reporting lately Zoom fatigue. Why should we care about the quality of video conferencing experiences?

There's probably a lot of reasons why meetings on Zoom and other video conferencing systems can be more tiring for a lot of people than meeting in person tends to be. And of course, some of that might just be that we're not only talking to other people, we're also potentially dealing with technological problems and things like that.

But also, we as humans really evolve to deal with other people in person. And the video conferencing experience is different from dealing with people in person in a number of ways, and that can make it more challenging, and probably that is some of why it's more tiring.

I see. So, what do you see as the biggest problems with video conferencing today?

OK. Well, in a high level, when we're in person, we make use of a lot of social cues to decide whether or not other people are paying attention, to decide how they're reacting to what we're saying, whether or not they like what we're saying or don't like what we're saying, we get a lot of social cues for whether or not someone wants to take turns in a conversation, whether they want to break in and talk.

And also for that matter, when we meet someone new, we're trying to decide whether or not we want to have a relationship with this person and whether or not we want to trust them. And it's not clear that we can do this in video conferencing in the same way and as well as we do in the real world.

I see. But you're a vision scientist. So why are you interested in this?

Well, because fundamentally, a lot of those social cues for them to work, visual perception needs to work. Before you can ask whether or not someone can make use of the social cues, you've got to ask whether or not they can actually perceive what they need to perceive in order to make the judgments that they need to make.

So we need to not only adequately perceive individuals, their facial expressions, and what they're doing and so on, but also, we need to be able to perceive groups of people as a whole. In a video conferencing we're often faced with a lot of people's faces, and there are a lot of fundamental questions in visual perception that are involved in whether or not that perception is successful, whether it's effortless, and therefore, enjoyable.

And what are some of the variables of video conferencing environments that complicate human vision research?

Well, one of the biggest issues is clutter. And we know that human visual perception is worse when there's a lot of visual clutter, and this is one of the main things that we study in my lab. And that clutter comes not just from having a lot of things in the display-- we tend to think of clutter as having too much stuff on our desk. And it's not just the number of things, it also can come from variability in the display.

So it's not just the number of people on the video conferencing call, it's variation in their appearance, in the backgrounds of their videos, in what direction they're facing their heads, in their facial expressions, in the way they're moving, and so on. That variability can make processing that scene more difficult. And of course, other things can matter too, like the spacing of the participants, and the layout of the videos, and so on.

In video calls with a lot of participants, there can be a gap in the quality between cameras, and lighting, or sound quality. Do these things make a difference?

Yes, absolutely. That's another way in which participant videos can vary a lot, and that causes clutter. And we know, for instance, that how well you can perceive someone's facial expressions depends a lot on how high contrast their features of their face are, which has to do with how well they're lit. And in turn, that perception of facial expressions can be really important, because that can have an effect on whether or not you trust someone, whether or not you want to establish a relationship with them and so on.

How about that? Who knew there was so much to lighting?

Yeah. [LAUGH]

There's a lot going on when you have multiple people on a video conference call. Are we, as humans, capable of making sense of all that?

Yeah. Interestingly enough, there's some previous work on this and the results are definitely hopeful. So we know, for instance, that perhaps it's not so surprising that we're very good at judging characteristics of a single phase, like the facial expression, but it turns out people are also good at judging properties of the whole set of faces.

You can get the average emotions, like on average, are these people in the audience enjoying the talk or something like that, you can get a quick sense of the average direction people are facing and things like that. And people can do that very quickly, relatively effortlessly, even though there's lots of people to process. However, we need a lot more research on this topic because this issue of visual clutter is very important.

Vision scientists tend to study these kind of questions in very homogeneous displays, meaning you've got a lot of very similar faces, they're against a completely blank background, there's none of these sources of variability that I'm talking about. And a lot of what my lab does is try to extend our understanding of human vision to the kinds of variability that you see in the real world, or in real applications.

Are there any things people can do to improve their video conferencing experience that are supported by research?

Well, unfortunately, there's not a lot that individuals can do on their own. Of course if you have the resources, improving your internet bandwidth will help, it'll improve your experience as well as other people's experience of you. Similarly, making sure the camera has a good view of your face, lighting yourself well so that you're nice and well-lit from the front and there's lots of good contrast on your face. That sort of stuff will all improve how people perceive you.

Other than that, probably the best thing you could do is to not look at your own video. There have been some suggestions that seeing your own face is very distracting. You're spending part of your time thinking about how you look and how you're responding, and that could add to Zoom fatigue. Essentially, you're multitasking, and multitasking is tiring.

But fundamentally, a lot of the issues require design changes and technological solutions. And they're not so much something an individual can implement.

What are some of the ways this proposed research could benefit video conferencing applications?

Well, video conferencing still seems like a pretty competitive space. And so I think the main benefit is that if companies can change their designs or their technology in a way that's inspired by things like human vision, people might enjoy using their product more and perhaps they'll even keep using it when we're not in a pandemic.

Now this question might seem a little strange, but is the use of video conferencing or can the use of video conferencing have an impact on climate change?

So first of all, I should throw in the caveat that of course, this is not my topic per se, but it does seem like that is possible, although the story is a bit mixed. So it's fairly clear that video conferencing is sure a lot better than getting on a plane and flying to a meeting across the country or something like that, but it's not necessarily better for climate change than just heading into the local office.

Are there other industries that could benefit from improvements in computer-mediated human vision research, like telehealth or education?

Well, there's tons of stuff where they could benefit from human vision. Of course, there's an immediate connection just between if you can make video conferencing better then that changes how we think about online learning, it potentially changes how we think about things like telehealth.

Although, you meeting with your doctor has different vision issues than you meeting with a large group of people because it tends to be one-on-one and so there's not as much of the issue of trying to process a whole bunch of faces all at the same time.

Is there any other research you're currently excited to share?

Sure. Well, more generally, my lab studies human vision and how it relates to applications. And a lot of our interest is in what information is available to a user at a glance at a display. And what information is available at a glance has a lot to do with how you process information across your field of view, meaning not just where you're pointing your eyes, but all over the visual field in front of you.

To the extent that we can understand the processing at a glance, we can understand a lot about what tasks will be easy and what tasks will be hard for the user, we understand a lot about usability and the user's experience. Because if you can do the task at a glance, the task is going to be faster and easier because you don't have to move your eyes, you don't have to piece together information from one glance to another.

In the last decade or so, my lab has made significant progress in understanding visual processing across the field of view and what you can see at a glance. And we're very excited to move some of that new basic science into applications in how to better design user interfaces and things like that, based on what we know about human vision now.

You talked about trust and building trust with video. What makes people trust someone else? Is smiling beneficial?

Yeah, I think that the knowledge about trust comes from a mix of psychology and probably the human-computer interaction field. And it's a variety of things, but some of it, I'd say probably a fair amount of it is how well you can predict someone else's responses. That has a lot to do with trust. Also in human-robot interaction, a lot of it has to do with prediction, it has to do how well you trust the robot also.

And so it's whether someone's responding the way you think they should respond to the situation that you're both facing is a lot of how you decide whether or not you trust another person.

Great. Well, Ruth, thank you very much for your time today, we appreciate it.

Thank you. Nice to talk to you.

[MUSIC PLAYING]

If you're interested in learning more about the CSAIL Alliance program in the latest research at CSAIL, please visit our website at cap.csail.mit.edu. And listen to our podcast series on Spotify, Apple Music, or wherever you listen to your podcasts. Tune in next month for a brand new edition of the *CSAIL Alliances* podcast, and stay ahead of the curve.

[MUSIC PLAYING]