# Audio file

[OII Podcast - Mohsen and Cameron.WAV](https://unioxfordnexus-my.sharepoint.com/personal/inet0394_ox_ac_uk/Documents/Transcribed%20Files/OII%20Podcast%20-%20Mohsen%20and%20Cameron.WAV)

# Transcript

00:00:04 Veena McCoole

Hello and welcome to the OII podcast brought to you by the Oxford Internet Institute, part of the University of Oxford. In this podcast, we explore issues and developments in the digital world that matter to us all.

00:00:17 Veena McCoole

I'm your host, Veena McCoole, and today we're joined by two experts in the field of misinformation and polarisation, Professor Mohsen Mosleh and Cameron Martel.

00:00:26 Veena McCoole

Mohsen is associate professor of social data science here at the OII. His research focuses on how misinformation and disinformation spread on social media and how ties are formed on social network.

00:00:38 Veena McCoole

Cameron is a PhD candidate at MIT Sloan School of Management and an incoming assistant professor at Johns Hopkins Carey Business School, focusing on why people believe and share misinformation, what forces shape the online social networks through which misinformation spread.

00:00:54 Veena McCoole

And which content moderation interventions are effective for improving online information quality?

00:01:00 Veena McCoole

Mohsen and Cameron thank you so much for joining us today.

00:01:03 Mohsen Mosleh

Thank you for having us.

00:01:04 Cameron Martel

Yeah. Thank you so much.

00:01:05 Veena McCoole

So to kick things off, can you paint us a picture of the current state of misinformation on social media today, including maybe some of the tactics used to spread disinformation online?

00:01:16 Mohsen Mosleh

Sure. So given the true information accounts for substantial portion of information shared on social media ecosystem.

00:01:24 Mohsen Mosleh

There is empirical evidence, including work by myself, across multiple platforms, that low quality information and misinformation are significantly shared more and receive more engagement than true information, and that.

00:01:39 Mohsen Mosleh

When you control for various features of the content, such as political lean and the characteristics of the user who initially created content and part of this is due to the tactics used by malicious actors such as creating emotionally.

00:01:55 Mohsen Mosleh

Arose out misleading content and content that are novel and engaging, yet not true.

00:02:01 Cameron Martel

Yeah. And also add on to what to what Mohsen said here that.

00:02:08 Cameron Martel

This type of exposure and also sharing of false and low quality content is also very.

00:02:15 Cameron Martel

Heterogeneous and you can vary a lot by different subpopulations. There's evidence, particularly in the US and amongst Americans, that this can vary a lot by political orientation and my age, for instance. So certain pockets of people online can be exposed to and engage with misinformation at much higher rates.

00:02:33 Cameron Martel

Than perhaps on average across everyone.

00:02:36 Veena McCoole

OK, so can you explain why people believe and share misinformation and low quality content?

00:02:43 Mohsen Mosleh

Sure, most people share misinformation, either because they are not paying attention or they are confused and mistakenly think the information is true. Only a small number of people, contrary to what many people think, is that they would share.

00:02:59 Mohsen Mosleh

Misinformation purposefully to create chaos.

00:03:02 Mohsen Mosleh

In some of our work through a survey experiment, we found that like around 50% of people share misinformation.

00:03:08 Mohsen Mosleh

10 because of inattention on about 35% was caused by confusion and only 15% and less was intentional. These findings have been confirmed by other studies in different settings.

00:03:24 Mohsen Mosleh

There are generally 2 main explanations for why people struggle to tell the truth from falsehood. 1 is political motivated reasoning. That is, people may believe and share information that aligns with their political identity.

00:03:41 Mohsen Mosleh

And dismiss information that doesn't.

00:03:44 Mohsen Mosleh

Because they are motivated to protect their identity, another one is classical reasoning. That is, the ability to tell the truth from falsehood depends on analytical thinking. People who reflect and think more critically are better. Identifying false information, regardless of whether it matches their political beliefs.

00:04:04 Mohsen Mosleh

Recent studies.

00:04:06 Mohsen Mosleh

Strongly support the classical reasoning explanation. That is, people who think more carefully are less likely to believe or share misinformation, no matter what their political leanings are.

00:04:18 Cameron Martel

Yeah. And I think something that is important to to add on is to that people being confused, confused and mistaken account of why people might sometimes believe in share misinformation. This doesn't necessarily mean that people are really gullible.

00:04:35 Cameron Martel

Or being duped necessarily, but can stem from really rational and reasonable heuristics or kind of rules of thumb based on what people are typically exposed to in their information or social media environment?

00:04:48 Cameron Martel

So for instance, some of my early work looks at if people who engage in more intuition or emotion believe in share misinformation more, and we find that they do, and that experimentally manipulating and inducing people to rely on emotion more.

00:05:03 Cameron Martel

Increases their belief in false, but not in true news.

00:05:07 Cameron Martel

Later, I kind of wanted to reevaluate some of that research and look at why that might be the case. Where might that rule of thumb or heuristic be coming from that makes people who are using more of that intuition and emotion actually believe the false?

00:05:20 Cameron Martel

So it's more and so with my colleague Reed Orchinik and several other collaborators, we varied the type of content that people saw in their feeds. If they saw mostly true or mostly false content, with the idea being that maybe this intuition, this heuristic that people are using.

00:05:36 Cameron Martel

Is stemming from what they're typically exposed to.

00:05:39 Cameron Martel

We find that when people mostly see true content, which is the case for most people in their online information environments, people are more likely to take false stuff as true because they're learning a base rate.

00:05:50 Cameron Martel

Rule of thumb that most of what they're going to see is true, so follow that you. If you're acting quickly. If you're just using that heuristic, would mistake false is true now. On the flip side, if you're seeing mostly false stuff, you actually become overly sceptical and you're more likely to believe true stuff.

00:06:05 Cameron Martel

To be false and make that kind of mistake. So I think even within that kind of classical reasoning account, people are using pretty reasonable and adaptive strategies for how they're engaging with information.

00:06:18 Cameron Martel

Even within that, this can lead to why people are still able to believe and share falsehoods within these different online environments.

00:06:27 Veena McCoole

OK, so it's clear our brains can easily be tricked into believing and reinforcing false information on occasion. I'm curious, how does this connect to the rise of polarisation on social media? Can you explain how maybe these digital environments shape our views?

00:06:43 Cameron Martel

Yeah. Yeah, so I think.

00:06:47 Cameron Martel

A lot of how people engage online is largely reflective of their offline environments as well and.

00:06:53 Cameron Martel

The main difference I I would ascribe to what digital online environments provide is perhaps like a change in the the magnitude or quantity of the type of homopholus or similar experiences that people might be exposed to. There is a lot of offline for example.

00:07:10 Cameron Martel

Segregation of people by political beliefs and a lot of this is reflected online as well. That research from most and myself shows, and I think it's this scale at which people can kind of make these decisions as to how they're shaping their environments that can then.

00:07:26 Cameron Martel

Kind of spillover and lead to continual effects into what they're exposed to and who they're exposed to online.

00:07:33 Veena McCoole

So misinformation isn't just about facts, it's about how people interact with them. And it may seem self-evident. But how does this relate to the idea of polarisation? What kind of real world impacts arise when people increasingly identify with, say, their political tribes online?

00:07:50 Cameron Martel

Yeah. So I think my work with motion suggests that.

00:07:54 Cameron Martel

People largely form online networks because of partisanship per SE as a causal factor, so it's not only that online algorithms are necessarily recommending people to a friend, partisans, Co partisans and not friend counter partisans who they disagree with, but people actually have these preferences much like they would offline as well.

00:08:15 Cameron Martel

And other work, a virus suggests that this is due in part because of people wanting to connect with like others, and also have a disc reference for connecting with people who do not share our same partisanship.

00:08:28 Cameron Martel

This also includes like type prevention, like blocking people are way more likely to block counterpartisans they are to block Co partisans, or politically neutral accounts.

00:08:38 Cameron Martel

And in terms of how this effects kind of polarisation, we find that this is largely symmetrical in the US amongst Democrats, who are left-leaning Republicans on the right, such that people are similarly likely to follow back Co partisans regardless of their partisanship. We're reducing some interesting differences in our field work.

00:09:00 Cameron Martel

Is that Democrats are more likely to block Republicans and vice versa, and we have some evidence suggesting that this might stem from Democrats having a greater preference for blocking people who perhaps post low quality or more toxic content, which is reflected in what some of our studies have found these Republican profiles to be sharing.

00:09:19 Cameron Martel

And I know most of them have some other work speaking to similar things as well.

00:09:24 Mohsen Mosleh

So sure, I mean, one of the thing that we found and this is a paper that came out a couple of months ago, we found that like right leaning users are more likely to share misinformation on the platform. That is even though like like ring users are more likely to be suspended.

00:09:41 Mohsen Mosleh

On the platform, this is not easily. We cannot easily draw the conclusion that this action is driven by the politics itself, but it's more kind of like related to the kind of like behaviour like online behaviour of these users in terms of like.

00:09:56 Mohsen Mosleh

Like problematic content such as misinformation.

00:10:00 Veena McCoole

So let's move on to what's being done to combat misinformation. Can myths and disinformation be effectively corrected or debunked once it spread?

00:10:12 Cameron Martel

Yes, and a lot of my work looks at different types of, uh, content moderation interventions that tech platforms can and in some cases have applied to trying to reduce the belief in and spread of falsehoods online.

00:10:26 Cameron Martel

Umm some of this work. I think it's helpful to think of it in terms of interventions that are content neutral. So it doesn't depend on a specific headline being evaluated and content specific ones that really target specific claims or posts.

00:10:41 Cameron Martel

So some of the content neutral approaches that myself and other people researching misinformation have found to be effective.

00:10:46 Cameron Martel

Are kind of digital literacy tips and education so kind of teaching people how what to look for in posts that might be low quality. So examining the source, examining different features of the headline.

00:11:01 Cameron Martel

Engaging in something called lateral readings. So trying to search across different sites to validate information and these can be delivered in the form of when you go online, maybe they're ten different tips that you might see before you start scrolling on Facebook.

00:11:15 Cameron Martel

That could be helpful other work by myself and from my collaborators David Rand and Gordon Pennycook and most in here as well, have looked at the efficacy of accuracy prompts. So essentially reminding people to think about and consider accuracy.

00:11:31 Cameron Martel

They go about trying to share different types of information, as most alluded to earlier, most people aren't purposely sharing falsehoods, but this can happen in perhaps if they're not thinking about accuracy when they're making sharing decisions. So once you remind someone to have accurate.

00:11:47 Cameron Martel

Top of mind.

00:11:48 Cameron Martel

Kind of close the gap between what they believe and what they actually wind up.

00:11:53 Cameron Martel

We find that in a recent adversarial collaboration with myself and some other collaborators across different research teams found that this can even be effective across the political spectrum for both those on the political right and left. When you give people these accuracy prompts, it can increase the quality of what people share online.

00:12:13 Cameron Martel

These are some of the more content neutral strategies. Content specific strategies can be in the form of Corrections coming after the fact, or a fact checker warning labels which can be applied directly on.

00:12:24 Cameron Martel

Posts, for instance, a recent paper of mine looks at the efficacy of fact checker warning labels on belief in and sharing of false news. We find that not only do these warning labels work on average at reducing the belief in and sharing of false posts, but they even work for people who say that they strongly distrust fact checkers, which is the source of where these come from.

00:12:44 Cameron Martel

And I think suggest that this type of warning label approach can be really effective. This warning strategy had been used by Facebook and Instagram previously for several years and recently it's been announced that they've are facing out this strategy.

00:12:58 Cameron Martel

Which I think will have important consequences for the information quality of these news environments.

00:13:05 Cameron Martel

Finally, one thing that I have done some recent work in lately is if we can use crowdsourcing and kind of collective intelligence strategies to scale up the ability to identify and act upon potential falsehoods or misleading content online.

00:13:20 Cameron Martel

One notable example of this is X's community notes, which started when it was called Bird Watch under Twitter, and essentially what this allows users to do is when they encounter a post that they think is false or misleads or misleading, they can classify it accordingly and write a free response note or a Fact Check.

00:13:37 Cameron Martel

Adding additional context as to what might be misleading, and then users can evaluate the helpfulness of those notes on other people's.

00:13:43 Cameron Martel

Posts and I think strategies like this can be really effective for helping scale up the ability for users to engage in content moderation on these platforms.

00:13:55 Veena McCoole

O to what extent are social media to blame? What is their responsibility and all of this?

00:14:01 Mohsen Mosleh

So these studies generally suggest that people share misinformation because they forget to think about how accurate the content is.

00:14:11 Mohsen Mosleh

Before they, you know, they before they make a sharing decision, the issue isn't necessarily that people are, you know, lazy thinkers in general. Instead, there is reason to believe that social media by design distracts us from the concept of accuracy by focusing our attention.

00:14:28 Mohsen Mosleh

On social factors, things like how many likes my post will get, what comments people would make on so on. So even people who under normal circumstances think very carefully they wind up mistakingly sharing misinformation while.

00:14:42 Mohsen Mosleh

Online, but also this means that social media platforms can change their design. Things like Cameron for example, mentioned to add to the, you know, social media features.

00:14:55 Mohsen Mosleh

Social media companies have the power to focus attention on accuracy and help immunise against misinformation. If they really choose so.

00:15:06 Cameron Martel

Yeah, I think there's nothing new about misinformation or falsehood or polarisation or any of these phenomena that we that we're talking about that suddenly social media has, like invented or brought about now. That said, I do think that there are, as Mosin mentioned, substantial differences in kind of how this information is portrayed. So the difference is.

00:15:26 Cameron Martel

Distractions or combination of the the social and then the news media. Part of when accuracy really matters. If you're engaging with actual news headlines with when it matters a lot less. Perhaps if you're engaging in more entertainment posts or just trying to connect with friends online.

00:15:40 Cameron Martel

I think the combination of these can make it really.

00:15:43 Cameron Martel

Tricky for users and even for platforms for designing the best way to make sure that users are engaged with high quality content in a way that makes sure that they're informed and not misinformed from what they're from what they're reading.

00:15:56 Veena McCoole

Thank you both for those insights.

00:15:59 Veena McCoole

Now I'm interested to hear what role governments are playing in combating misinformation and what more should be done to tackle this problem at a policy level.

00:16:08 Mohsen Mosleh

So, I mean, governments can introduce laws to make social media companies accountable for their content shared on their platform, as well as increasing transparency and providing data access to researchers. We have seen some initial efforts in that front, for example through the Digital Service act in the EU, but have certainly not enough in terms of.

00:16:29 Mohsen Mosleh

You know, researchers and independent, you know, researchers have access to the, you know, data on the platform at scale.

00:16:37 Mohsen Mosleh

Governments can also empower fact checking organisations and collab, collaborate with them to promote true information, and reduce the spread of false content.

00:16:47 Cameron Martel

Yeah, in the US, these policy decisions have really largely been made by the technology platforms themselves and those in charge of them.

00:16:57 Cameron Martel

As we've seen recently, Meta and Mark Zuckerberg announced massive changes to how Facebook and Instagram would be engaging with their content moderation practises ending in the US. Their partnership with independent professional fact checkers.

00:17:11 Cameron Martel

In evaluating content that would be sent to them, and then if it was false, be typically labelled and demoted or downranked in people's feed so that they'd see it less regularly. Instead, they will be moving as they've announced to something more similar to exes. Community notes approach without any fact checking, similar to how.

00:17:31 Cameron Martel

Elon Musk, when he acquired Twitter, got rid of a lot of their trust and safety team as an existent.

00:17:36 Cameron Martel

Previously.

00:17:38 Cameron Martel

I think that as researchers.

00:17:42 Cameron Martel

Two really important things to consider here are one. If we have the data and access available to the content moderation approaches, and perhaps algorithms used, for instance in a Community note system for what notes are actually being surfaced.

00:17:56 Cameron Martel

To continually evaluate if these content moderations are working, what users are being exposed to and what is happening on these platforms, essentially. And then second, rather than treating these different interventions as trade-offs with one another, necessarily saying that.

00:18:12 Cameron Martel

We should have user moderation instead of professional moderation. Figuring out the best ways to combine these different approaches in order to create the best experience for users and create the best information environment for constituents.

00:18:27 Cameron Martel

And hopefully with this type of transparency and being able to assess what's happening on platforms, we can continue to to have good evidence as to what content moderation works and also examine what users.

00:18:43 Cameron Martel

Want themselves?

00:18:43 Cameron Martel

Happy using on platform.

00:18:46 Veena McCoole

So you've both discussed and published recent research on these topics, but I'm curious to hear where are the big research gaps? You know, if you could weigh the magic wand or indeed expedite the research process, what would you want to know more about and where do you think more attention is needed?

00:19:03 Mohsen Mosleh

Sure. I mean one big gap is that most of the research work in this area, I've been focusing on you know.

00:19:11 Mohsen Mosleh

A single social media platforms, namely Twitter, now called X and a bit of research on other major platforms like Facebook for example. But none of these platforms by no means are representative of social media ecosystem at large.

00:19:27 Mohsen Mosleh

So it is important to see how patterns we find in one single platform is universal versus like a specific to a specific platform.

00:19:38 Cameron Martel

Yeah, two kind of big broad picture themes that I'm interested in and this would be a case across this, this multi platform landscape that that Mosin talks about here are harm and content moderation, scalability. So on the harm side, I think something that can be overlooked.

00:19:56 Cameron Martel

Sometimes in in research on social media is really thinking about the outcomes and like user behaviour, particularly offline, that we care about and that we want to perhaps protect consumers.

00:20:10 Cameron Martel

From so one example would be a recent paper from my colleague Jennifer Allen that came out in science.

00:20:16 Cameron Martel

Examines what type of content on Facebook was actually harmful in terms of increasing scepticism towards COVID-19 vaccines.

00:20:26 Cameron Martel

And there are kind of two main findings that she and her colleagues found in this paper. 1 is that conditional on actually being exposed to and viewing certain.

00:20:36 Cameron Martel

Posts outright false posts about the COVID-19 vaccine decrease people's perceptions of safety and willingness to get the vaccine more than true, but like potentially misleading posts. So that would be, for instance.

00:20:51 Cameron Martel

A headline that says a healthy doctor died two weeks after getting a COVID vaccine. So if you saw it, then the false post would be have a worse impact. However, what Jenny finds is that if you look at who actually saw these posts on Facebook during the pandemic.

00:21:06 Cameron Martel

That it's this lighter type of content. It's this misleading stuff. That way more people.

00:21:10 Cameron Martel

City and so kind of taking that into account, what the effect of actually seeing the content, it's the true book misleading stuff that had a way bigger aggregate impact on decreasing people's intentions to get a vaccine, vaccinate their kids and so on. I think this is really important because one, I think it speaks in part to.

00:21:28 Cameron Martel

Moderation efforts on the fall stuff probably being pretty good at the time. Now those systems are changing, so we'll have to evaluate how things proceed in the future. But two, I think it speaks to the fact that.

00:21:40 Cameron Martel

It's another type of online content that is perhaps causing this societal harm that we should think about. Different ways to do content moderation on. I don't think, for instance, this true content deserves a warning label. Perhaps that says false as it is not, but thinking about different ways that we might be able to do content moderation and make our feeds better for.

00:21:59 Cameron Martel

The people that use them.

00:22:01 Cameron Martel

And then the second point on scalability related to the work on looking at community notes and other crowd sourced approaches, I think figuring out how to have sustainable.

00:22:12 Cameron Martel

Perceived as legitimate mechanisms by which users can kind of curate the environments that they want, and not necessarily that the platforms want them to have in these environments is really important. So perhaps figuring out different panels or sustainable ways such as like on subreddits or Facebook groups where users can really have a hand in how they want.

00:22:33 Cameron Martel

Platforms to look like is one thing that I'm very interested in looking at more going forward as well.

00:22:39 Veena McCoole

It sounds like there's still a lot to find out.

00:22:43 Veena McCoole

Finally, I'm curious about what all of this means for the future. With the rise of generative AI, deep fakes ever evolving social media algorithms, what does the future hold for misinformation and disinformation on social media, and what changes do you see coming?

00:22:58

Sure.

00:22:58 Mohsen Mosleh

I mean, with a recent, you know, development of generative AI, you will and we'll have already been seeing that like multimodal misinformation, which is more realistic and.

00:23:11 Mohsen Mosleh

Us were targeted audience and at larger scale.

00:23:14 Mohsen Mosleh

But on the other hand, though, AI can also be used to fight misinformation as well. At scale. For example, recent work by our colleagues Tom Costello, Gord Pennycuke and Dave ran at MIT, showing that conversation with AI can improve.

00:23:30 Mohsen Mosleh

Beliefs against conspiratorial beliefs, and it can scale up fact checking.

00:23:37 Cameron Martel

You know, one thing that I'm perhaps most concerned about with new advancements in general AI large language models is specifically that scale aspect that Mohsen mentioned in so far as I think there's going to be a lot more content that is this.

00:23:56 Cameron Martel

Gen AI creative artificial posts and images, texts that people will see.

00:24:04 Cameron Martel

As well, and I think that it really lowers the cost and level of entry for people, perhaps with bad intentions to kind of flood different environments with this type of content. I think that's one reason why figuring out content moderation approaches that work and work in different places is really important.

00:24:23 Cameron Martel

For instance, 1 like kind of subcategory of kind of false and misleading posts could be things that are outright frauds and scams that can cause people online to lose a lot of money, whether it be.

00:24:35 Cameron Martel

On Facebook marketplace and retail scams, or with other false postings, different types of click, crypto engagement, and so on.

00:24:43 Cameron Martel

And so I am worried, not necessarily, that llm's are going to make this content necessarily more persuasive or better, but mainly that they'll just be so much more of it because it's going to be so much easier and cheaper to produce. So again, figuring out how to match the scalability of this type of content.

00:25:02 Cameron Martel

Creation. That's bad with a scalability of moderation that can kind of quality effects and keep feeds having good and reasonable and not dangerous content, I think is something that's going to be a struggle going forward.

00:25:16 Cameron Martel

Or tech platforms and researchers like myself to figure out how to do best.

00:25:20 Veena McCoole

OK. So it sounds like two things AI can actually be supportive in the fight against misinformation, but equally really compound its effects. Thank you both so much for your time today. It's been fascinating to explore the world of misinformation and polarisation with you, and thank you to our listeners for tuning into this episode of The OII Podcast.

00:25:38 Veena McCoole

If you've enjoyed it, please leave us a review and share the link to this episode with your network. We'll be back soon with more conversations about the digital world. Take care.