

WHO CAN WE TRUST ONLINE?



SEASON 5, EPISODE 7
PRESENTED BY



S05E07: Who Can We Trust Online?

Richard: Hello and welcome to Crossing Channels. I'm Richard Westcott. Who can we trust online? That's the subject of the latest in our collaboration between the Bennett School of Public Policy at the University of Cambridge and the Institute for Advanced Study in Toulouse at the Toulouse School of Economics. As ever, we're going to draw on the interdisciplinary strengths of both institutions to explore a complex issue. How do people decide what to trust online? What role do platforms, algorithms, and AI systems play in shaping what we see? And what kinds of governance, awareness, and public institutions are needed if we want a healthier information environment? To explore these issues today, we have Rachel Adams from Cambridge University. Rachel, start us off. What does your research focus on?

Rachel: My research really is about understanding the impacts of artificial intelligence on global inequality. And I'm interested in this in kind of both directions. One, how can AI genuinely expand and support human flourishing in every corner of the world. And two, how is artificial intelligence undermining the conditions that make global justice possible in the first place. And I'm particularly interested and have done a lot of work around the impact of AI on people living in extreme poverty.

Richard: Joining us from the Toulouse School of Economics, we have Tiziana Assenza. Tiziana, remind us of your main research interests.

Tiziana: So I'm a macroeconomist and I'm interested in studying how individuals form expectations about main economic variables and how these in turn affects their economic decisions and behaviour. Starting from this, lately I have been more more interested in studying how misinformation can distort economic decisions. And I really start from what individual households, firms or investors believe and do all the way up to the business cycle

fluctuations. So it's really on both, I look at it from both lenses of the micro and the macro impact. And in turn, then how making people aware of their blind spots can somehow have an impact on the way in which they behave.

Richard: Online information is central to how we learn about the world, but how can we filter out misinformation, fake news. Algorithmic amplification and now AI videos that look real. Now I'm fascinated in this topic, especially because I'm a former journalist, so actually the reality of what you see online is very, very important to me. Rachel, your work looks at AI governance, global inequality, and the political economy of emerging technologies. Is online misinformation mainly down to false content, or is there a deeper problem with the systems that determine what information reaches people in the first place?

Rachel: I think it's important to remember that the infrastructure, that these messages, content, claims are being circulated on are owned by a very small number of firms. So there is a very powerful kind of concentration of power and influence within the infrastructure on which the information environment of today's age now sits. And so it's not just a question of whether there's a validity in a particular claim that might be within a particular piece of content. The bigger questions are around why is that content or that claim reaching that person in that form at that scale at that particular point in time? And these are all questions that are determined by the kind of platforms and the infrastructure that are the kind of channels through which this information travels. So I think it is a much deeper systemic problem. And we can get into some of the kind of governance implications or considerations at a later point.

Richard: Yeah, because the BBC, we were just very heavily governed by rules and you had to follow those broadcast rules. Silly question really, it's a bit basic, but I'm assuming there are no actual rules.

Rachel: There are some rules, like in the UK we have the Online Safety Act. In the European Union there is digital services laws that pertain to platforms. But a large part of my work is reviewing the kind of equality that exists globally around how strongly these platforms and AI more generally is governed. And so what we find is that in Europe and countries like the UK, there are quite effective governance regimes quite effective in comparison to absolutely zero. They're not entirely effective. Whereas in other parts of the world, there's very few hard regulations. So whatever kind of policies or frameworks that might exist are non-binding in nature. They have very little effect. We're kind of relying on how far a platform might choose to self-regulate. And so there's enormous disparity between how protected an individual in the UK is compared to an individual in Ethiopia. And we've seen the fallout of this.

Richard: Tiziana, your research on fake news looks at how well people identify false informations and how they perceive their own ability to recognise what's true. We all think we're good at filtering it out, but actually are we? What do we know about why people are vulnerable to misinformation, even when they may believe they're careful or they're well informed?

Tiziana: Well, there is a large research both in psychology and political science that studies exactly this, so why we got fooled by misinformation. And we can think of three mechanisms that dominate in this respect. So the first is cognitive. So research in this area essentially has

shown that detecting misinformation requires slow, deliberate thinking. And when you think about most of us, we simply don't engage this kind of attention when we are scrolling on the social media, on the media in general. And this can somehow explain why we fall for misinformation. Then there is a second mechanism that is mostly motivational. So here the researchers, they refer to what they call identity protective cognition. So the idea is that when you are in front of a piece of information that is threatening your political or social identity, you have the tendency not to evaluate it neutrally. Instinctively what you do is to defend yourself and the group to which you belong to. And then there is a third mechanism that is mostly emotional. So there is research that has been shown that on what it was Twitter, false news spread about six times faster than true news. And this is largely explained by the fact that they trigger stronger emotional reactions like anger, fear or surprise. So people have a tendency to share much more false news because they are driven by this emotional aspect. This means that essentially we know quite a lot about why people fall for misinformation, while what is less understood is how aware they are about their own vulnerability to misinformation. And if we make them aware of their vulnerability, whether this is gonna actually change their behavior. And this is exactly what we track in our paper. Is this going to have an impact on their beliefs and in turn on their economic behavior? What we find is a paradox because when you ask people how good they think they are in spotting misinformation, we find that more than eight out of 10 reply that they are good. When you ask them what they think about the average American, only three out of 10 say that the average American is good. So you see that essentially they recognize that the problem exists, but it's not their problem, it's the problem of somebody else. And they behave as if they are disconnected from the impact of the behavior that this somebody else can have on their own life.

Richard: When you're talking to people, are they surprised at what they're getting wrong? How do they react?

Tiziana: Well, actually, what is nice is that the people that we reach with our study are the ones that you want to reach, the overconfident ones. What we find is that these are the ones that react mostly to the information about their true susceptibility to misinformation and in the right direction. So they revise downwards their beliefs about their ability and they increase their willingness to pay for protection against misinformation. So somehow we are acting at the metacognition level. So we are not going to tell people, I'm gonna teach you how to spot misinformation. We are just telling them it exists. You recognize yourself that it exists. We are gonna give you your numbers, your statistics, not just some statistics. And let's see how that impacts.

Richard: Well let's talk about some of the resolutions to this as you've brought that up. Now, Rachel, you mentioned Ethiopia, for example. How do you protect people from being misled? I mean, ironically, the way you reach lots of people is the internet. So the the very tool that you're t you're telling them to be aware of is the tool that w to reach them. Are there any solutions out there? Rachel, we'll start with you.

Rachel: I mean I think it's really a complicated issue when we're talking about how do we regulate and govern this kind of thing? Because I work a lot with African governments and what we've seen in the past is some kind of content appearing online being amplified by various kinds of tools, digital tools, and it being content that a government doesn't want to

see online and so their response has been internet shutdowns. So we really don't want to be creating the kinds of legal mechanisms that would enable a more authoritarian leaning government to do some kind of blanket shutdown and squash freedom of expression. So that's the kind of extreme end of the issue that we're dealing with. The kind of question about how literate are people to discerning and being skeptical about the kind of content that they are viewing and accessing is incredibly important and is one of the main ways that we want to be able to kind of support a critical democratic information environment. But we are not in a good place globally. So a think tank that I ran out of South Africa, did a public perception survey with over 3000 South Africans from all walks of life in all different languages to ask them about their understanding about artificial intelligence, where they're getting their information from, how they use digital platforms, etc. And this is important for the misinformation, disinformation piece because so much misinformation, disinformation is either created or amplified with generative AI technologies in particular. And one third of South Africans had never heard of AI. Another third had heard of it but had no idea what it was. So we're looking at two thirds of South Africans don't know about artificial intelligence. Most of them are online. And then the majority of people, if they do know about AI or what they do know about AI, is received through digitally mediated platforms. And so they're learning about AI on these platforms. They're learning about how AI can distort what they're receiving on these platforms. So that is really, really worrying. And I think the other consideration, just to go back to the point about Ethiopia, because it's important to really kind of properly understand how differently these problems manifest and turn up in different parts of the world, is that platforms have a responsibility to curate to an extent any content online that is harmful in some way. So in America where it's like two main languages, Spanish and English, they have thousands of moderators that are checking online content to make sure it's not content that would be harmful and taking that down. In Ethiopia, for example, where there are 85 languages being spoken online, content moderation, a report by Amnesty International showed that their content moderation abilities were only in four languages and that they had no human content moderation abilities in the lead up to the 2021 election in Ethiopia. And there's been really serious conflict in the Tigray region in Ethiopia for a number of years, much of which has been inflamed by the kind of content that's appearing online. We think that these platforms have AI systems that are able to read inflammatory content, but these AI systems do not work in the 85 Ethiopian languages. They will maybe work in two, perhaps three of them, and not well enough to be able to pick up dialect or slang or the nuances of idiom in which very, very harmful content can appear. So I think you've kind of got three levels there. You've got where does the government sit and how far leaning are they towards democracy authoritarianism? So how far do we trust our government to govern truth in a way that is in the public interest and preserving of democratic values? Then you've got how literate are populations? How far do they? Understand these things, how far do they understand how AI is changing these things? And then you have what country are we looking at and how far is the social media platform in that country? How far are they putting in place safeguards to ensure that the content that appears online isn't harmful? From there, we can then look at an attenuated governance regime, particular to those different circumstances. But as you can see, we can't put in place blanket laws that are going to work for everyone everywhere.

Richard: That is a fairly bleak picture with some pretty stark numbers in there, isn't it? But you know, that's the reality, isn't it? That's the reality of the world. Tiziana, is there any way we can overcome some of these clearly huge issues?

Tiziana: So let me just briefly explain how we built our experiment because this will help me to bring to where I want to get. Essentially what we did is to design a really simple experiment in which people in our experiment don't just read about misinformation, but they are facing their blind spot directly. Okay, so their vulnerability. So we asked them at the beginning how good they think they are in spotting misinformation and we also ask them to make an hypothetical financial decision. And they have to allocate the budget, including how much they would like to spend to protect themselves from the harming of misinformation. So then they are confronted to the evaluation of 20 real headlines that have been fact-checked by professional fact-checkers as being true or false. And halfway through these tests, there is a treatment group that is informed about their actual score while the control group sees nothing. So they are not informed about the actual score. And then we ask again participants what they think is their own score after they've been exposed to the treatment and how they would allocate again the budget. Like that, we can measure whether indeed there is any implication of informing people about their true vulnerability in terms of both belief updating and economic decisions that in our case is proxied by this willingness to pay for protection against misinformation. And what we find is that indeed the ones that are overconfident are the ones that are revising more strongly their beliefs. And this revision, and this is the really, I think, the most important part of our experiment, translates directly into economic behavior. So we have that every one point drop in their self-assessment leads to about \$6 extra investment in misinformation protection. So this brings me to what we could do. I think that we should in the long run build an apparatus that collects all the tools that we have out there. Because right now we know that fact checking and media literacy are needed and they work. So my argument is not at all against them. What I'm trying to propose is a tool that can be a complement with respect to this type of actions. Why? Because when you think about fact checking or media literacy, they are operating at the object level. Okay, so they teach you about the threat what manipulation looks like, how you can spot a misleading headline. But what they are not able to do is to tell you how well you are actually judging. And that's the meta level on which our intervention acts. So while the object level is strengthening the toolkit, what we want to do with our intervention is to calibrate the toolkit user. I see two practical implications of this. The first is that it's somehow self-targeting because essentially the biggest impact lands on exactly those people that you need to target, that are the overconfident one. And you don't need to identify them in advance. And secondly, it works even when platforms are not willing to cooperate or when regulation is low because in principle it does not depend on anyone else acting first but it can operate in parallel. What we show is that the lever is overconfidence so if we are able to address it then belief and behavior, if you believe in our results, will follow.

Richard: What role do you think schools should play here? Because we're talking predominantly about adults, but actually children are being completely washed over with information online, aren't they? And how do we make sure that the children know, probably through school, what's actually going on?

Rachel: I have four children and so I'm really worried about the content that they are being exposed to and their inability to know what's true or false or what's being doctored in some way and my general sense is it's not their responsibility to know these things. And yes, I think it would be important for schools to do more to raise awareness. We know that schools do things around online bullying, and I think this is kind of an extension of that kind of issue.

But I think it might be going more in the direction of bans for particular ages. I know there's a lot of issues around age verification systems for different apps and children being able to get round them, but children are the next generation and I think so much of what makes disinformation and misinformation so worrying is how far it pushes us towards more populist politics. So I think it is something that we need to be taking very seriously.

Tiziana: Yeah, in the medium long run, we should really invest in teaching kids first of all about the fact that the problem exists. And then we should give media literacy on misinformation. mean, we have the example of Finland that has been doing it now for about almost 20 years, starting from the kindergarten. Now, regarding the access to the social media, I mean, in general, I don't like banning. I mean, what I like is to teach people how to use a tool. My fear is the following. In the best-case scenario, when you ban something, it can be the case that at the end you have no effects. But there can be a worse scenario in which you have bad effects because if you ban the access to specific social media then it can be the case that kids go towards even more dangerous places in the internet. So at the end you get much negative impact of the policy than a neutral one.

Rachel: I just want to add on a piece about the children because there is a level of interventions that's around teaching people to identify and understand what they see. But there is also a set of responsibilities that governments hold about ensuring there are mechanisms available for redress and remedy where harm actually does occur. And we know that teens are particularly vulnerable to manipulative content online, which is a kind of sister to misinformation and disinformation. We've just conducted a global study called the Global Index on Responsible AI, which assesses how far 135 governments around the world are taking measures to protect people's rights in the age of AI. What we're seeing is very, very few countries around the world have any kinds of laws in place to ensure that there are institutions that people can go to, mechanisms for redress and remedy and justice if a child is harmed with AI and online. So that's something we really, really need to address.

Richard: In your book, *The New Empire of AI*, you examine how AI can deepen global inequalities and reshape power. You've been talking about that quite a bit. How do we make sure that debates about misinformation and AI governance don't simply reflect the concerns of richer countries?

Rachel: One of the things we need to be thinking about is to strengthen the capacity of independent institutions around the world. think we often talk about platforms and people and citizens and governments, but actually independent institutions think tanks as well play a really, really important role. So in Africa, we have a group called Africa Check that is independent, has existed for many years and fact checks information that is not being fact checked anywhere else. So they're doing really, really valuable work and they are the kind of group that we want to be supporting more of. The we here is probably international development, aid and philanthropy, particularly low resource governments are not going to be funding civil society. Now the other one is independent state institutions. So in South Africa, for example, we have something called the Competitions Commission and they have recently put a huge fine on Google because what it did was prioritise international news over local news in its search engine. And often local news is closer to the ground truth and offers slightly more diverse, nuanced perspectives to reporting locally. And so the competitions commission fined Google for this kind of obvious favouritism of international news sources

over local media. But it also showed how important these state institutions are for addressing systemic structural issues that will not be addressed by an individual, but also for supporting kind of public debate around some of these issues. And so yeah, I think that there's much more we can do there to support these kinds of institutions to do work locally and then that feeds up into global discussions.

Richard: Tiziana, so what does this mean for democracy and public debate, especially when citizens are making decisions in environments where false or misleading information can spread very quickly?

Tiziana: Well, here the stakes are concrete. I mean, when you think about the individual level. It's clear. I mean, this can materialize with financial scams, dangerous health decisions. It can distort the voting choices. And this is something that research has been analyzing, investigating a lot. But what is even more worrisome is that these consequences don't stop here. I mean, I'm a macroeconomist. So as much as I'm interested in understanding individual behavior, my final objective is then to eventually and how this individual behaviour transmits and translates into aggregate outcomes and how this works at the macroeconomic level. So that is why in another piece of research that I've been conducting, what we do is to look at how misinformation that has technology-related content, so for example fake news about innovation, AI or digital infrastructure, can have an impact on the macroeconomy on the business cycle fluctuations. And here the logic is simple. You have to think of this technological content misinformation as introducing noise in the system, in the information environment. So if the information environment is more noisy, then this makes it harder for households, firms, and markets in general to have an accurate read of the state of the economy. This implies that uncertainty increases, but together with the increase in uncertainty, we will also observe an increase in disagreement. People are starting to form different pictures of the same environment in which they are acting. And if this is the case, so if you have a situation in which households, firms and investors are not able anymore to form consistent expectations about the economy, then this can have non-negligible negative impact on the economy as a whole. So you can observe a falling consumption, a slowdown in hiring and therefore an increase in unemployment and a post-mortem of investments when you think about the behaviour of firms. And what we find in our research is exactly this path. But what is really striking is also the magnitude of this impact.

Because let me give you an example to make it comparable with something that we know. So when you study the macroeconomic impact of a one percentage point increase of the monetary policy rate by the central bank, so this is what we define an aggressive monetary tightening, usually this takes one year and a half to fully transmit to the real economy. What we find in our paper is that technological content misinformation produces a comparable effect on industrial production, but in one month. And differently from monetary policy, in this case, there is not a central bank that can respond in real time to this negative impact to the economy. I think that what is also surprising is that it's not any kind of fake news. It's really technology related fake news. So it's really the supply side ones because we did investigate also other types of contents like fiscal deficit, other types of economic contents, and we didn't find any impact. I mean, the impact comes really from technology related misinformation. From a in terms of policy implications, that essentially we should be able to monitor these type of topics and the same way in which we monitor other macroeconomic risk indicators. I mean, this is something that we are already doing. So we should integrate

like an early warning system for those type of misinformation that is economically sensitive. And this is somehow like a natural extension that we could think of when we think of the risk dashboards that central banks or financial ministers are already using. So I think that we should indeed treat misinformation as what it is, as systemic risk.

Richard: And do you think governments, politicians get that? Do they understand the economic dangers?

Tiziana: I think that they are somehow afraid now more and more of it. They are becoming more and more aware that misinformation is not just a threat to democracy as it has been studied so far but can have economic implications that are not at all negligible. What I think it's not clear so far is that we shouldn't act only at the citizen level, but we should also somehow figure out a system that is able to act also at the more aggregate level.

Richard: Okay, we always like to finish by looking ahead into the future. So if we want an online information environment that people can trust, what is the one change you would each prioritize over the next decade? So it could be technology design, public policy, education, any of those things, anything else that you think of, but what's the one change you would prioritize and why, Rachel?

Rachel: The fundamental issue at play is that the information infrastructure globally, the incentives that underlie that are fundamentally misaligned with democratic values. They're about profit seeking and so they're about exploitation. So I think the kind of ideal would be that these information ecosystems are designed and governed in the public interest and that these companies are much more kind of public benefit organizations. I'm not sure that that's politically feasible. And so I would say that we need much more accountability, harder accountability and more transparency about how these systems work, what kind of information and content platforms are choosing to amplify, to recommend and why people are accessing them. I think if we had more understanding around that and harder rules to govern that, that would be a really important place to start.

Richard: Tiziana, where would you focus?

Tiziana: So what I would prioritize over the next decade is just bigger than a single tool. What I envisage is a shift in how we treat the metacognitive calibration, so the people awareness of their own informational limits. And we should think of it as a piece of public infrastructure. I mean, when you think about our current status, we already accept that the state is producing and protecting certain informational goods. Think about official statistics, public health communication or also about financial literacy programs or central bank communication. So here what we are missing, I think, calibration. So giving people accurate and personalized feedback about how well they are in processing the information environment in which they live. So in the short run, major platforms should be willing to offer periodic and private personalised feedbacks on the accuracy of individuals. And this could be helped by a regulatory requirement, this could fit naturally within the European Digital Service Act. You don't need to change the law. The law is already there. Then when you think about the medium run, this would mean essentially to integrate these metacognitive training into schools. And this goes back to your question about what do we do with kids? So it's not just about teaching them what fake news looks like, but also to give these kids

structured and repeated experience about the fact that they can be wrong about it and how they can eventually take action to protect them about it. And then finally, in the longer run, this means to treat the information environment as a macroeconomic variable. So something that the economic policy institutions, starting from central banks and to fiscal authorities, this should monitor as they do already with inflation expectations or financial stability indicators. And this is what the evidence from our experiment shows that this kind of intervention is able to change not only belief of the individuals but also their economic behaviour. I think that we shouldn't ask to any institution to adjudicate what is true in real time because this is somehow politically unmanageable. But most probably what we should do is to give people the tools that equip them to calibrate their own judgement repeatedly and at a low cost.

Richard: Well that's all we've got time for for this fascinating episode. Thanks to Rachel Adams from Cambridge University and Tiziana Assenza from the Toulouse School of Economics for joining us. And thank you for listening to Crossing Channels. We'd love to hear what you think. Your reviews help us shape future episodes and make it easier for new listeners to find the show. If you enjoyed this one, why not explore some of our earlier episodes too?