

# Theme 5: Disinformation, Epistemic Vices & Online Harm

Thursday, 6 May 2021

|             |  |
|-------------|--|
| 13:55-14:00 | Welcome  |
| 14:00-15:00 | <b>Alessandra Tanesini:</b> Arrogance and Anger on Social Networking Site (1 pm BST)<br><i>10 min break</i>                    |
| 15:10-16:10 | <b>Quassim Cassam:</b> Disinformation, Narratives and Radicalization (2.10 pm BST)<br><i>10 min break</i>                      |
| 16:20-17:20 | <b>Tatjana Scheffler, Veronika Solopova, Mihaela Popa-Wyatt:</b> The Telegram chronicles of online harm<br><i>10 min break</i> |
| 17:30-18:30 | <b>Kevin Zollman:</b> Homophily, polarization, and epistemic performance: a simple model (11:30 am EDT)<br><i>10 min break</i> |
| 18:40-19:40 | <b>Cailin O'Connor:</b> Retraction in Scientific Network (9:40 am PDT)   |
| 19:40-20:15 | Optional round-table discussion  |

Friday, 7 May 2021

|             |  |
|-------------|--|
| 13:55-14:00 | Welcome  |
| 14:00-15:00 | <b>Stephan Lewandrosky:</b> The Knowledge Dementors (1 pm BST)<br><i>10 min break</i>  |
| 15:10-16:10 | <b>Anastasia Kozyreva:</b> Psychology of disinformation and cognitive tools against online manipulation<br><i>10 min break</i> |
| 16:20-17:20 | <b>Eric Beerbohm:</b> Gaslighting Citizens (10.20 am EDT)<br><i>10 min break</i>   |
| 17:30-18:30 | <b>Michael Lynch:</b> Social Media, Conspiracy and Bald-Faced Lies (11:30 am EDT)<br><i>10 min break</i>                       |
| 18:40-19:40 | <b>Ray Drainville &amp; Jennifer Saul:</b> Visual and Linguistic Dogwhistles (12:40 pm EDT)                                    |
| 19:40-20:15 | Optional round-table discussion  |

## ABSTRACTS: DAY 1

### **Arrogance and Anger on Social Networking Site**

ALESSANDRA TANESINI (Cardiff University )

2:00 pm CET — 1:00 pm BST

Anger dominates debates in the public sphere. Discussions especially on social media quickly and frequently become shouting matches whose participants are not afraid to express their contempt for each other. My focus is especially on two types of anger that I label respectively arrogant and resistant. The first is the characteristic defensive response of those who unwarrantedly arrogate special privileges for themselves. The second is often a source of insight and a form of moral address. I detail some discursive manifestations of this two types of anger. I show that arrogant anger is responsible for attempts to intimidate and humiliate others with whom one disagrees. Whilst resistant anger can be intimidating, it is also essential in communicating moral demands. I conclude the talk by demonstrating that calls for civility, especially on-line, always risk perpetrating injustices.

### **Disinformation, Narratives and Radicalization**

QUASSIM CASSAM (University of Warwick)

3:10 pm CET — 2:10 pm BST

In this talk I'll examine the role of narratives in online radicalization. I will introduce and explore the notion of a radicalization narrative and focus on the question whether it is helpful to think of such narratives as consisting of disinformation. I will argue that counter-radicalization requires the development of effective counter-narratives that are deep, credible, relevant, and resonate with their target audiences. I will conclude by examining the role of truth in anti-extremist counter-narratives.

### **The Telegram chronicles of online harm**

TATJANA SCHEFFLER, VERONIKA SOLOPOVA, MIHAELA POPA-WYATT (Ruhr-Universität Bochum, Freie Universität Berlin, ZAS Berlin)

4:20 pm CET

Harmful and dangerous language is frequent in social media, in particular in spaces which are considered anonymous and/or allow free participation. In this talk we focus on a Telegram channel populated by followers of Donald Trump, in order to identify the ways in which harmful language is used to create a specific narrative in a group of mostly like-minded discussants. We provide an extended taxonomy of harmful language, including hate speech, direct insults, and also indirect ways of poisoning online discourse, such as divisive speech and the glorification of violence. We apply this taxonomy to a large portion of the corpus. Our data gives empirical evidence for harmful speech such as in/out-group divisive language and the use of codes within certain communities which have not often been investigated before. We discuss some of the challenges that the Telegram data set poses for automatic methods of harmful speech detection.

### **Homophily, polarization, and epistemic performance: a simple model**

KEVIN ZOLLMAN: (Carnegie Melon University)

5.30 pm CET — 11:30 am EDT

Many have suggested that our tendency to interact with those who have similar views (belief homophily) contributes to polarization. I construct a simple model of belief-based polarization that shows this, and allows us to understand how information, misinformation, and homophily interact to create the conditions for polarization. In addition, I explore some of the epistemic consequences of homophily and polarization. I will connect the model to two other famous simple models of belief changes based on "opinion dynamics."

### **Retraction in Scientific Network**

CAILIN O'CONNOR (UC Irvine)  
6:40 pm CET — 9:40 am PDT

Sometimes retracted or thoroughly refuted scientific information is used and propagated long after it is understood to be misleading. Likewise, sometimes retracted news items spread and persist, even after it has been publicly established that they are false. In this paper, we use agent-based models of epistemic networks to explore the dynamics of retraction. In particular, we focus on why false beliefs might persist, even in the face of retraction. We find that, paradoxically, sometimes a delay in retraction may make it more relevant. We also find that the network location of retraction is very important in determining its impact.

## ABSTRACTS: DAY 2

### **The Knowledge Dementors**

STEPHAN LEWANDROSKY (University of Bristol)  
2:00 pm CET — 1 pm BST

We are said to live in a "post-truth" era in which "fake news" has replaced real information, denial has compromised science, and the ontology of knowledge and truth has taken on a relativist element. I argue that to defend evidence-based reasoning and knowledge against those attacks, we must understand the strategies by which the post-truth world is driven forward. I depart from the premise that the post-truth era did not arise spontaneously but is the result of a highly effective political movement that deploys a large number of rhetorical strategies. I focus on three strategies: The deployment of conspiracy theories, the use of "micro-targeting" and "bots" online, and agenda-setting by attentional diversion. I present evidence for the existence of each strategy and its impact, and how it might be countered.

### **Psychology of disinformation and cognitive tools against online manipulation**

ANASTASIA KOZYREVA (Max-Planck Institut, Berlin)  
3:10 pm CET

In this talk, I will discuss psychological underpinnings of why disinformation spreads online and how to address this imbalance with interventions that empower Internet users to gain some control over their digital environments, for instance by boosting their information literacy and their cognitive resistance to manipulation.

### **Gaslighting Citizens**

ERIC BEERBOHM (Harvard University)

4:20 pm CET — 10.20 am EDT

Gaslighting, as an interpersonal wrong, brings its victims to doubt the sources of their evidence. This paper holds that political gaslighting, by leading citizens to hold beliefs disconnected from the available evidence, poses a distinctive threat to democratic politics. But holding “audacious beliefs” — beliefs that are ahead of the evidence — can serve as a core ingredient for democratic movements. This creates a dilemma for citizens, who must choose between two kinds of evidential policies. How can they protect themselves from the gaslighting without rendering themselves insusceptible to the mobilizing efforts central to democratic politics? Citizens, then, face a standing challenge: to remain open to the bully pulpit while vigilant against the epistemic bullying that characterizes gaslighting.

### **Social Media, Conspiracy and Bald-Faced Lies**

MICHAEL LYNCH (University of Connecticut)

5:30 pm CET — 11.30 am EDT

Social media is widely used to push not only outright conspiracy theories but what me might call “political bald-faced lies”, or the postings of obviously false propositions. Not all such bald-faced lies are made in the context of conspiracy; nor are they plausible efforts to deceive. So what is the purpose of such postings, particularly by those in power? I’ll argue the point is to express that power and to undermine the social-epistemic rules and practices so important to democracy.

### **Visual and Linguistic Dogwhistles**

RAY DRAINVILLE & JENNIFER SAUL (University of Waterloo)

6:40 pm CET — 12.40 pm EDT

There is now a burgeoning literature in philosophy of language on dogwhistles—a particular sort of coded utterance, common in (but not limited to) politics. This paper explores the neglected but extremely important visual dimension of dogwhistles.