Selected Papers of #AoIR2023:
The 24th Annual Conference of the
Association of Internet Researchers
Philadelphia, PA, USA / 18-21 Oct 2023

MISOGYNY, SURVIVORSHIP, AND BELIEVABILITY ON DIGITAL PLATFORMS: EMERGING TECHNIQUES OF ABUSE, RADICALIZATION, AND RESISTANCE

Sarah Banet-Weiser
University of Pennsylvania

Kathryn Claire Higgins
Goldsmiths, University of London

Nelanthi Hewa
University of Toronto

Debbie Ging
Dublin City University

Catherine Baker
Dublin City University

Maja Brandt Andreasen
Dublin City University

Panel Rationale

On 18th May 2022, in an opinion piece for *The New York Times*, columnist Michelle Goldberg declared “the death of #MeToo” (Goldberg, 2022). The immediate context for her declaration was the hypermediated defamation trial involving actor Amber Heard and her ex-husband Johnny Depp, which had, at the time of Goldberg’s column, just about swallowed the internet whole. Though the verdict of the trial was still pending, the public verdict was resounding: it seemed, per *Vice*, that “the entire internet [was] Team Johnny Depp” (Zoledziowski, 2022). Meanwhile, Heard was being subjected to what Goldberg describes as “industrial-scale” bullying and misogynistic abuse. That Depp

Suggested Citation (APA): Banet-Weiser, S., Higgins, K.C., Hewa, N., Ging, D., Baker, C. & Brandt Andreasen, M. (2023, October). *Misogyny, Survivorship, and Believability on Digital Platforms: Emerging Techniques of Abuse, Radicalization, and Resistance*. Panel presented at AoIR2023: The 24th Annual Conference of the Association of Internet Researchers. Philadelphia, PA, USA: AoIR. Retrieved from http://spir.aoir.org.

had a well-documented history of violent and abusive behavior seemed to matter little to those rallied under #JusticeForJohnnyDepp. He was a victim of unchecked and disingenuous feminism – and the feminists were finally getting what they deserved.

The papers in this panel take the claim of “the death of #MeToo” seriously and wrestle with its potential implications. Broadening the lens beyond this one case, they examine case studies and data from the United States, Australia, the United Kingdom, and Ireland in order to evaluate the current state of play in the online push-and-pull between feminist speech about gender-based violence and its attendant misogynistic backlashes.

It is certainly nothing new to suggest that social media imply certain vulnerabilities to harassment, abuse, and disbelief that are explicitly gendered and raced, as feminist and anti-racist scholars of digital media have long documented (see Noble, 2018; Banet-Weiser, 2018; Steele, 2021). Similarly, there is a long tradition of scholarship on the “digital manosphere” that has foregrounded social media as a space of networking, capacity-building, and radicalization for misogynistic movements (see Gotell & Dutton, 2016; Ging, 2019; Johanssen, 2021). In 2018, Sarah Banet-Weiser theorized this terrain as a “funhouse mirror” of (digitally mediated) popular feminism and popular misogyny, arguing that the key resources and strategies of the former were being steadily appropriated and distorted to fortify the backlash politics of the latter. Today, we’re confronted with an uncomfortable and urgent question: *is the backlash winning?*

Each paper in this panel approaches this fraught terrain with a different set of questions and different methodological approaches, but with the same concerns in mind.

Nelanthi Hewa draws from semi-structured interviews to examine how journalists reporting on sexual violence and gender-based abuse are orientated by the logics of digital platforms. Hewa’s analysis unpacks how the incitement to visibility and exposure that characterizes social media (Gray, 2013) is compounding with the incitement to transparency that characterizes journalism. They conclude that the imperative to have “nothing to hide” has important implications for both the journalism of sexual assault and the online and offline vulnerabilities of sexual assault survivors.

Building on this commentary, Sarah Banet-Weiser & Kathryn Claire Higgins shift the lens from visibility to believability, analyzing what happens to and with the different forms of digital evidence that survivors are incited to share publicly when they make sexual assault allegations. Drawing from a conjunctural critique, Banet-Weiser & Higgins propose that the gendered and racialized politics of doubt has been “digitized” in ways that tend to fortify, rather than disrupt, prevailing patterns of suspicion and distrust. They propose that an emerging strategy of online networked misogyny is to exploit the vigilance towards signs of inauthenticity and manipulation that characterizes contemporary internet culture in order to keep sexual violence allegations in a state of permanent “irresolvability.”

Debbie Ging, Catherine Baker & Maja Brandt Andreasen provide context for these analyses by examining one of the key phenomena shaping the push-and-pull between feminism and misogyny online: the radicalization of boys and men into misogynistic ideology. Using a set of experimental social media accounts, Ging, Baker & Brandt Andreasen test the proposal that social media platforms contain "radicalization pathways" that nudge users towards increasingly more radical far-right and misogynistic content, combining a thematic analysis of content with close monitoring of algorithmic recommendations and trajectories.

References

Bailey, M. (2014) More on the origins of Misogynoir. *Moyazb*, April 27. https://moyazb.tumblr.com/post/84048113369 /more-on-the-origin-of-misogynoir.

Bailey, M. (2021). *Misogynoir Transformed: Black Women's Digital Resistance.* New York University Press.

Ging, D. (2019). *Gender Hate Online*: *Understanding the New Anti-Feminism*. Palgrave Macmillan.

Goldberg, M. (2022). Opinion: Amber Heard and the death of #MeToo. *New York Times*, May 18. https://www.nytimes .com/2022/05/18/opinion/amber-heard-metoo.html.

Gotell, L., and Dutton, E. (2016). Sexual violence in the "manosphere": Antifeminist men's rights discourses on rape. *International Journal for Crime, Justice and Social Democracy*, 5(2), 65–80. https://doi.org/10.5204/ijcjsd.v5i2 .310.

Gray, H. (2013). Subject(ed) to recognition. *American Quarterly*, 65(4), 771–798. http://www.jstor.org/stable /43822990.

Johanssen, J. (2021). *Fantasy, Online Misogyny, and the Manosphere: Male Bodies of Dis/Inhibition*. Routledge.

Steele, C. (2021). *Digital Black Feminism.* New York University Press.

Zoledziowski, A. (2022). Why does it seem like the entire internet is team Johnny Depp? *Vice*, April 25. https:// www.vice.com/en/article/4aw93j/justice-for-johnny-depp -internet-comments.

Platformed Survivorship: Digital Platforms and Digital Exposure

Nelanthi Hewa

Research Question/Issue

In 2000, an editorial by an anonymous journalist describing being groped by Prime Minister Justin Trudeau was published in a local paper; in 2018, a photo of the editorial was tweeted by an unaffiliated political pundit with the hashtag #metoo. In this paper, I bring Sara Ahmed's (2006) theory of queer orientations into conversation with scholarly work on digital techno-capitalism and publicity online to explore how digital media orient sexual violence survivors and the journalists who report on them toward publicity and exposure. I take up the groping accusation against Justin Trudeau in the summer of 2018 as a refraction point to follow different rays of light as they dance on the wall: the internet as a public space, the framework of visibility that dominates sexual violence coverage online, and the interplay between legacy media outlets (classically thought of as slower, more ponderous, less reactive) and digital-first news outlets. I explore the orientation towards publicity/virality of digital media in conjunction with the chase for virality by journalists and the competition between outlets, as what began as an offline anonymous editorial moved through Twitter, a gossip magazine, *BuzzFeed*, and finally Canadian legacy media outlets like the CBC and *the National Post*.

The journey of the groping allegation—from anonymous editorial to tweet to ammunition used by conservative outlets, to BuzzFeed—exemplifies the process by which sexual violence becomes news as well as how those standards of newsworthiness have shifted. Indeed, that the allegation was treated as gossip but also as potential news, or incipient news, exemplifies the complexity of what journalists in our interviews term the "lowering of the bar" when it comes to allegations of sexual violence. While standards around verification were not entirely gone, there was nevertheless a feeling that this was news, rather than "simply gossip." Though *Frank* magazine*, BuzzFeed* and the *CBC* all have different standards of what passes as news and covered the story differently, all of these outlets nevertheless acted as though reporting on the allegation was necessary.

Methodology/Critical Framework

In *Queer Phenomenology: Orientations, Objects, Others*, Sara Ahmed (2006) attends to the way orientations — sexual orientations, but also orientations as ways of being in the world more generally — come to be, and come to be felt on the body. She argues that, when we are oriented *towards* a table (for example), we are also oriented *away* from what is behind the table, so that to queer one's orientation would be to ask: who can sit at the head of the table? Who is under it? Who keeps it clean, and who is bereft of a seat? I take up Ahmed's phenomenology of queer orientations as a starting point for a materialist media theory to ask similar questions of the media and platforms at which we sit, since a table, like Twitter, is also a platform (Singh & Banet-Weiser, 2022). When one uses media, how does it orient us? Or, when media call to us, how are we expected to respond? A user that turns towards a machine's hail might be its ideal user, one who fits comfortably into its contours. In being hailed by media, however, what comes to be the "background" or the private?

I bring this theoretical framework into conversation with a discourse analysis of news stories and data from semi-structured interviews I conducted with Canadian journalists who cover sexual violence. Building off the work of feminist scholars who have argued that women have been constructed as not credible and discounted as liars (Jordan, 2004), I move to look at the role of digital media in challenging and narrowing feminist politics of speaking out. From the moment you press "enter," digital life and the internet are replete with metaphors of space. "Cyberspace," windows," and "entering" websites all direct us when we go online to think of the internet as a bounded space. Moreover, to go online is to be interpolated as an active and empowered “user,” in control of the interface, situated within the screen, and moving actively through internet “space” (White, 2014, p. 1). Yet one cannot enter the interface, delete unwanted search results, fully delete one’s data, or even know where, exactly, one’s data is going in these “leaky networks” (Chun, 2016). As scholars have increasingly argued, not all users are archived by the internet’s algorithms of oppression (Noble, 2018) in the same way. The metaphor of (public) space employed by ICT companies, journalists, and users alike is central to how we conceptualize and interact with the digital, yet it fundamentally papers over the realities of what it is to be online. Though the metaphor of the open public space pulls us to believe that to be out and visible forever is a good, even a necessity, for survivors, standards of proof and believability have hardly shifted to become more accommodating of the complexities of sexual violence.

Findings/Conclusions/Relevance

Sexual violence stories’ online circulation illuminates the particularly insidious aspects of the visibility economy because stories of sexual violence all rest on a particular kind of evidence. Evidence of the crime is fundamentally entwined with evidence of the victim’s believability—their transparency—since these crimes, especially at the level of national coverage, are often taken as “he-said-she-said” stories. Competition between news outlets pushes journalists to cover stories quickly, an experience captured by an anonymous journalist I spoke with who, describing their work at a digital news outlet, said there was an attitude with stories that “[if] it was all over social media, it was like, let’s get this out as soon as possible because other news outlets are covering it.” Feminist ethics, I argue, are ultimately incommensurable with the attention economy of both the internet and contemporary news publications. In its orientation towards transparency, digital media demand that we have nothing to hide, a demand with particular implications for sexual violence news coverage. This is the paradox at the heart of the “internet as public space” metaphor, particularly as it relates to sexual violence: survivors are called to “rise up” and “speak out” in the visible platform of the internet, with the logic that to be out and visible (forever) is good for survivors. These material conditions exacerbate the limitations of a feminist politics of speaking out that “both supports survivor stories and requires survivors to tell these stories and to tell them in specific ways” (Serisier, 2018, p. 11). Sexual violence complicates—or perhaps enlivens—Sarah Banet-Weiser’s (2018) framework of visibility and popular feminism because *what* sexual violence survivors are asked to make visible is unique from other instances of violence or feminist public speech. Because survivors are often unable to make visible their lack of consent, it is survivors’ personhood and version of the facts that are debated as they are forced to run a “societal gauntlet of doubt, dismissal, or outright disbelief” (Epstein & Goodman, 2019). The only way for survivors to prove not simply their innocence but their *lack of guilt*—that is, that they are *not lying*—is to go always-more public, showing more and becoming hypervisible themselves in order to provide “proof.”

References

Ahmed, S. (2006). *Queer Phenomenology: Orientations, Objects, Others*. Duke University Press.

Banet-Weiser, S. (2018). *Empowered: Popular Feminism and Popular Misogyny*. Duke University Press.

Chun, W. H. K. (2016). *Updating to Remain the Same: Habitual New Media*. MIT Press.

Epstein, D., & Goodman, L. A. (2019). Discounting women: Doubting domestic violence survivors' credibility and dismissing their experiences. *University of Pennsylvania Law Review*, *167*(2), 399–462.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

Serisier, T. (2018). *Speaking Out: Feminism, Rape and Narrative Politics*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-98669-2

Singh, R., & Banet-Weiser, S. (2022). Sky High: Platforms and the Feminist Politics of Visibility. In S. Sharma & R. Singh (Eds.), *Re-understanding media: Feminist extensions of Marshall McLuhan*. Duke University Press.

White, M. (2014). *The Body and the Screen: Theories of Internet Spectatorship*. MIT Press.

Digital Evidence and Digitized Doubt After #MeToo

Sarah Banet-Weiser & Kathryn Claire Higgins

One of the most concrete changes brought about by the #MeToo movement is that it created a new public appetite for stories about sexual violence and other forms of gender-based harassment and assault. This appetite has been readily seized upon by Hollywood, the press, and through a renewed investment in digital media – especially, social media platforms – as a space where women and marginalized subjects can speak, gain visibility, and access public solidarity (see Gilmore, 2023; Jackson et al, 2020; Boyle, 2019; Durham, 2021). It has also spurred the growth of a new market for anti-sexual violence products and services – apps, wearable technologies, and other digital devices that promise that if only women can furnish more and better evidence of their assaults (photographs, videos, screenshots, and other 'corroborating' digital artifacts) then they will prevail in bids for believability, both in the court of public opinion and potentially in courts of law. However, at the core of this narrative are struggles about how, whether, and when different forms of digital evidence ought to bolster believability—especially now that such evidence can be freely circulated online, and highly public bids for belief made without arbitration or intervention by the state.

In this paper, we test the optimism of much #MeToo scholarship and of the emerging market for anti-sexual violence by tracking what actually happens to and with these digital artifacts in struggles over the "truth" of sexual violence allegations taking place on social media platforms. More precisely, we draw on conjunctural analysis (Hall, 1988; Clarke, 2014) to engage a close analysis of three key cases of disputed believability: in the US, the hypermediated defamation trial between actors Johnny Depp and Amber Heard; in the UK, the rape allegations levelled against Prince Andrew, a member of the British royal family, by trafficking survivor Virginia Giuffre; and in Australia, the rape allegation levelled against former attorney general Christian Porter. All three cases involved heated debates online over how, whether, and when different kinds of digital artifacts – or, what Adgebuyi (2021) calls "receipts" – should be allocated evidentiary value in struggles over the believability of sexual assault and gender-based abuse.

Our starting point for this analysis is the proposed analytic of a digitally mediated "economy of believability" (Banet-Weiser & Higgins, 2023) as a way of conceptualizing how different kinds of labor and resources come together on social media platforms to negotiate the "the truth" of sexual violence. Against a historical backdrop in which women and other marginalized people have been constructed as doubtful subjects *par excellence* and routinely doubted and disbelieved when they speak about their lives (Gilmore, 2017), we propose that there are new forms of labor designed to counteract this marginal positioning in the economy of believability through the production and circulation of ever-proliferating new forms of "proof."

For this reason, our analysis unpacks how contemporary struggles over the "truth" of sexual violence allegations are presently subject to what we term "the digitization of doubt." Broadly, this phrase captures how the different kinds of subject construction, labor, and performance that are implicated in struggles over believability are being increasingly routed through digital platforms and technologies, and so (re)shaped by and through their logics. When it comes to how the believability of sexual assault is negotiated online, our analysis of the Depp/Heard, Prince Andrew, and Christian Porter cases explores the gender and race politics framing the digitization of doubt across its three key core elements.

First, we examine the *new visibility and ease of circulation* of digital evidence. Afforming the arguments of critical race and technology scholars like Noble (2018), Brock (2021), and Benjamin (2019), we find that the apparent democratization of public speech and visibility that emerges with the digital landscape also means that it is ever easier for other users to cast doubt on marginalized individuals, as well as to attach clout to doubt by means of likes, shares, and comments. The "solidarity-building" capacities of social media, in other words, are an affordance without an inherent politics – and so, often, they work along pre-established lines of believability and cultural power. This is perhaps most immediately clear in the disparate size of Depp and Heard's respective online support-bases. On TikTok, #IStandWithAmberHeard had around 2.4 million views in early May 2022, compared with 6.8 *billion* views for #JusticeForJohnnyDepp (Siegel, 2022). The hypervisibility of the trial was also at the root of why so much misogynistic abuse emerged around it (see Hewa, 2021).

Second, we examine the *new forms and functions* of digital evidence. Here, we find that though survivors' "digital performances" of believability are now able to call upon more forms of proof, these forms are subject to ever-increasing levels of scrutiny and forensic deconstruction across YouTube channels, Twitter threads, online chat forums, and other locales in the digital manosphere, where images and texts can be easily cropped and chopped, removed from one context and placed in another, and wrapped in narratives that often fundamentally reshape their interpretation (Bock, 2021; Lampen, 2022). For this reason, digital forms of evidence are objects of heated suspicion, and occupy an uneasy status as "proof" in digital culture.

This hypervigilance towards signs of potential fakery or deceit fuses cultural anxieties about the untrustworthiness of women and queer people with anxieties about the capacities for deception and manipulation that are inherent to digital media, reinforcing both regardless of whether these artifacts are ever conclusively found to be fake (see Maddocks, 2020).

Third, building on what has been previously described as a "platformization of truth" (Cotter et al, 2022), we propose that we are also witnessing a *platformization of doubt.* Here, we find that all three of our cases chimed with the commercial logics of social media platforms in two important ways: first, in the attention-grabbing power of celebrity, and second, in the lucrativeness of controversy. Engagement-based revenue mechanisms do not discriminate based on the *reasons* for engagement, and so platformized struggles for the believability of sexual violence allegations are ultimately conducted under a commercial imperative to keep those allegations as contentious and outrage-worthy as possible (Young, 2020; Donovan et al., 2022).

We conclude by arguing that we need to radically shift the frame in which sexual violence is understood, from truth to believability. In a digital sphere wherein doubt is endemic, doubt saves accused men and sinks survivors. Ultimately, the digitization of doubt works to keep sexual violence allegations in a state of permanent irresolvability, so that accused men can retain public belief so long as any doubt remains. This is why, we argue, a preoccupation with whether allegations have been "proven above doubt" only gets us so far, as it sits in tension with contemporary practices of online public speech about sexual violence in ways that can't be ignored.

References

Adegbuyi, F. (2021, July 9). Is internet receipt culture our undoing? *Cybernaut.* https://every.to/cybernaut/is-internet-receipt-culture-our-undoing.

Banet-Weiser, S. & Higgins, K.C. (2023) *Believability: Sexual Violence, Media, and the Politics of Doubt*. Polity.

Benjamin, R. (2019). *Race after Technology: Abolitionist Tools for the New Jim Code*. Polity.

Bock, M. A. (2021). *Seeing Justice: Witnessing, Crime, and Punishment in Visual Media*. Oxford University Press.

Boyle, K. (2019). *#MeToo, Weinstein and Feminism*. Springer International Publishing.

Brock, A. (2020). *Distributed Blackness: African American Cybercultures.* New York University Press.

Clarke, J. (2014) Conjunctures, crises and cultures: Valuing Stuart Hall. *Focaal*, 70(1), 113–122.

Cotter, K., DeCook, J. R., and Kanthawala, S. (2022). Fact-checking the crisis: COVID-19, infodemics, and the platformization of truth. *Social Media + Society*, 8(1), 1–13. https://doi.org/10.1177/20563051211069048.

Donovan, J., Dreyfuss, E., and Friedberg, B. (2022). *Meme Wars: The Untold Story of the Online Battles Upending Democracy in America.* Bloomsbury.

Gilmore, L. (2023). *The #MeToo Effect: What Happens When We Believe Women*. Columbia University Press.

Gilmore, L. (2017). *Tainted Witness: Why We Doubt What Women Say About Their Lives.* Columbia University Press.

Hall, S. (1988). *The Hard Road to Renewal: Thatcherism and the Crisis of the Left*. Verso.

Hewa, N. (2020). The mouth of the internet, the eyes of the public: Sexual violence survivorship in an economy of visibility. *Feminist Media Studies*, 1–12*.* https://doi.org/10.1080/14680777.2021.1922483.

Jackson, S. J., Bailey, M., and Welles, B. F. (2020). *#HashtagActivism: Networks of Race and Gender Justice.* MIT Press.

Lampen, C. (2022). Which women do we choose to believe? *The Cut*, May 12. https://www.thecut.com/2022/05/why-do-so-many-people-think-amber-heard-is-lying.html.

Maddocks, S. (2020). “A deepfake porn plot intended to silence me”: Exploring continuities between pornographic and “political” deep fakes. *Porn Studies*, 7(4), 415–423. https://doi.org/10.1080/23268743.2020.1757499.

Siegel, T. (2022). Are Johnny and Amber stans for real? *Rolling Stone*, May 3. https://www.rollingstone.com/movies/movie -features/johnny-depp-amber-heard-fan-war-online-social -bots-1345208/.

Young, D. G. (2020) *Irony and Outrage: The Polarized Landscape of Language, Fear and Laughter in the United States*. Oxford University Press.

Maddocks, S. (2020). “A deepfake porn plot intended to silence me”: Exploring continuities between pornographic and “political” deep fakes. *Porn Studies*, 7(4), 415–423. https://doi.org/10.1080/23268743.2020.1757499.

Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press

Recommending Toxicity: Do the Algorithmic Recommender Functions of YouTube and TikTok Radicalize Boys and Men into Manosphere Ideology?

Debbie Ging, Catherine Baker & Maja Brandt Andreasen

It is widely accepted that platform recommender algorithms draw users into increasingly extreme content. This has become an issue of increasing concern in the context of manosphere influencer Andrew Tate’s TikTok videos racking up almost 12 billion views before his ban from the platform and subsequent arrest by Romanian police in January of this year. According to the radicalization hypothesis, channels in the intellectual dark web and the alt-lite serve as gateways to fringe far-right and other extreme ideologies.

However, there is considerable disagreement about this process of radicalisation in the scholarship on recommender algorithms. Papadamou et al. (2021) focused on quantifying the probability that a user will encounter an Incel-related video via YouTube’s recommendation algorithm. To do this, the researchers collected a dataset of YouTube videos shared on incel-related subreddits (6.5k) and a baseline dataset of random YouTube videos (5.7k). The researchers then developed a lexicon on incel-related terms to annotate videos as incel-related or non-incel-related. To examine the process of potential radicalisation on the site, researchers used the YouTube API to graph the recommendation algorithm output, using a random walker model to find the probability of users being algorithmically recommended Incel-related videos. They conclude that within five hops, when starting from a non-Incel-related video, this probability is 1 in 5.

In other work that supports the radicalization hypothesis, Ribiero et al. (2020) conducted a large-scale audit of user radicalization on YouTube by analyzing 330,925 videos posted on 349 channels, which they broadly classified into four types: Media, the Alt-lite, the Intellectual DarkWeb (I.D.W.), and the Alt-right. They processed 72M+ comments, and their analysis shows that the three channel types increasingly share the same user base; that users consistently migrate from milder to more extreme content; and that a large percentage of users who consume Alt-right content had consumed Alt-lite and I.D.W. content in the past. They also probed YouTube’s recommendation algorithm, looking at more than 2M video and channel recommendations and found that Alt-lite content was easily reachable from I.D.W. channels, while Alt-right videos were reachable only through channel recommendations.

In a systematic review to determine whether the YouTube recommender system facilitates pathways to problematic content such as extremist or radicalizing material, Yesilada, M. and Lewandowsky, S. (2021) found that most of the 23 included studies implicated the YouTube recommender system in facilitating pathways towards problematic content (i.e. conspiratorial content, anti-vaccination content, pseudoscientific content, content unsafe for children, Incel-related content, extremist content radicalizing content, and racist content). However, the review finds that many of the studies use random walk algorithms to determine the probability of encountering problematic content – which does not account for how the user personalisation and the interaction will influence the algorithm.

Contrary to these findings, Ledwich and Zaitzev (2019) contest the role of the YouTube algorithm in radicalisation. They categorized almost 800 political channels, differentiating between different political schemas in order to analyze the algorithm traffic flows out and between each group. After conducting a detailed analysis of recommendations received by each channel type, they refute the popular radicalization claims. On the contrary, their data suggests that YouTube's recommendation algorithm actively discourages viewers from visiting radicalizing or extremist content. Instead, the algorithm is shown to favor mainstream media and cable news content over independent YouTube channels with a slant towards left-leaning or politically neutral channels. Similarly, Munger and Phillips (2019) directly analyzed YouTube's recommendation algorithm and failed to find support for radicalization pathways. They maintain that the algorithm operates on a simple supply-and-demand principle and argue, rather than algorithms driving viewer preference and further radicalization, it is further radicalization external to YouTube which inspires content creators to produce more radicalized content.

A key issue with research to date in this area is that it relies predominantly on random walker models applied to graphs derived from API information or large computationally annotated datasets. Therefore, a prominent gap exists assessing the experience of "real" logged-in users traversing personalized algorithms based on viewing history. Due to the centrality of personalized algorithms recommendations to users' experience on social media sites, research addressing this gap is urgently needed. Moreover, the research to date has focused exclusively on YouTube and long-form video content. Given the recent surge in popularity of short video content (seen in the rise of TikTok and addition of YouTube Shorts), additional research is needed to explore how platform recommender algorithms function in these new format domains

The current paper reports on the findings of a study conducted in Ireland, which sought to document whether and how the recommender functions of YouTube and TikTok contribute to promoting misogynistic, anti-feminist and other extremist content to boys and young men. Using a research method derived from the Reset Australia project on YouTube manosphere radicalisation, and with support from the Reset Australia team, we set up experimental accounts on YouTube Shorts and TikTok to track the algorithmic recommendations and trajectories provided to fake male accounts on each of these platforms.

This short-term, qualitative study involved analyzing algorithmic recommendations and trajectories provided to 10 experimental accounts, 5 on YouTube Shorts and 5 on TikTok, as follows:
YouTube Shorts

- 2 boys under 18, who followed content at different points along the ideological spectrum, from more mainstream to extreme sources and influencers
- 2 young men over 18, who followed content at different points along the ideological spectrum, from more mainstream to extreme sources and influencers
- 1 blank control account that did not deliberately seek out or engage with any particular content, but instead followed the videos offered by YouTube Shorts' recommendations.

TikTok

- 2 boys under 18, who followed content at different points along the ideological spectrum, from more mainstream to extreme sources and influencers
- 2 young men over 18, who followed content at different points along the ideological spectrum, from more mainstream to extreme sources and influencers
- 1 blank control account that did not deliberately seek out or engage with any particular content, but instead followed the videos offered by YouTube's recommendations.

To monitor the process of algorithmic recommendation and trajectories, we created a codebook that details the signifiers of manosphere content. This includes e.g. known manosphere hashtags, known manosphere accounts, and manosphere topics. Once the codebook was established, we ran the experiment over the course of a week. Each account watched videos recommended by the TikTok and YouTube accounts respectively. The four accounts per platform (i.e. not the blank control accounts) "liked" and watched videos twice that were deemed manosphere content. They were gradually given "nudges" – e.g. watching 10 videos from known manosphere accounts – before returning to watching the suggested videos. This gradually led to an increase in manosphere content and also increasingly radical manosphere content. Videos identified as manosphere content were thematically analyzed – providing the research project with an array of themes and trends.

This paper reports on these themes, and details how the algorithms worked to introduce our fake accounts to increasingly radical manosphere content.

References

Ledwich, M. and Zaitsev, A., 2019. Algorithmic extremism: Examining YouTube's rabbit hole of radicalization. *arXiv preprint arXiv:1912.11211*.

Munger, K. and Phillips, J., 2019. A supply and demand framework for YouTube politics. *preprint*.

Papadamou, K., Zannettou, S., Blackburn, J., De Cristofaro, E., Stringhini, G., & Sirivianos, M. (2021). "How over is it?" Understanding the Incel Community on YouTube. Proceedings of the ACM on Human-Computer Interaction. https://doi.org/10.1145/3479556

Reset Australia. 2022. Algorithms as a weapon against women: How YouTube lures boys and young men into the ‘Manosphere’.

Ribeiro, M.H., Ottoni, R., West, R., Almeida, V.A. and Meira Jr, W., 2020, January. Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 131-141).

Yesilada, M. and Lewandowsky, S., 2021. A systematic review: The YouTube recommender system and pathways to problematic content.