PLATFORMS ON TRIAL: MAPPING THE FACEBOOK FILES/PAPERS

Matias Valderrama Barragán
Centre for Interdisciplinary Methodologies, University of Warwick

Introduction[1]

After the series of scandals around the US tech industry in the last decade, there has been a widespread backlash against Big Tech in what has come to be known as the Techlash (Foroohar, 2018). Multiple platform controversies have arisen over the interface designs, algorithms, and metrics of platforms, putting into question how they measure and intervene in our social interactions and their consequences to mental health, democracy and societies in general. Data leaks, whistleblowers, misinformation, election interference, and social experiments have raised alarm bells about the so-called toxic and unaccountable power of platforms and, more broadly, about the growing crisis of accountability in digital societies (Cooper et al., 2022; Khan, 2018; Marres, 2021; Suarez Estrada et al., 2022; Van Dijck et al., 2018).

The subject of this research is the question of the *platform harms* and the demand to hold platforms accountable. From a theoretical perspective informed by pragmatist and dramaturgical approaches (Boltanski, 2011; Goffman, 1986), societal harms are not objective conditions inherent to static technologies nor subjective definitions by individual humans. Instead, we need to look at the situations where the ways of interacting with these platforms are problematised and called to account for their configurations. Rather than attempting to prove platforms' harmful or toxic effects, this research aims to study how the connection between platforms and harms is made and unmade in the first place. Situated at the intersection of Digital Sociology, Media Studies, and Science and Technology Studies (STS), this research inquiry into *how actors make and unmake connections between social media platforms and societal harms across different settings*, in order to contribute a better understanding of how platforms are put on trial.

Methodology

Suggested Citation (APA): Valderrama Barragán, M. (2024, October). Platforms on trial: Mapping the Facebook Files/Papers. Paper presented at AoIR2024: The 25th Annual Conference of the Association of Internet Researchers. Sheffield, UK: AoIR. Retrieved from http://spir.aoir.org.

To address these questions, I am mapping a specific platform controversy: *the Facebook Files/Papers*. This controversy started in September 2021 when the Wall Street Journal began publishing a series of stories called the *Facebook Files* about the leak of internal documents of the tech company Facebook -now Meta (Horwitz, 2021). A few weeks later, the former Facebook employee Frances Haugen would reveal her identity as the whistleblower behind the leak on television and at a US congressional hearing, causing a major media uproar. Haugen denounced that her former employer was privileging profits over users' safety. By the end of October, more than a hundred news articles would be published based on Haugen's disclosures under the label of the *Facebook Papers* (Hendrix, 2021). Along with the huge media coverage, the internal documents would circulate through government agencies, congresses, courts, and advocacy groups, and being translated and reframed in different ways. Altogether, this is a fruitfully case to study platform controversies and exploring how various actors establish, demonstrate, and challenge the connection between platforms and societal harms across multiple settings.

Building on the tradition of controversy analysis (Jasanoff, 2019; Marres, 2015; Nelkin, 1992; Venturini & Munk, 2022), since October 2021 I have been *following the disclosures* across different media settings, combining digital and ethnographic methods in a mixed design methodology. Digital trace data collected through web scraping and API requests, are interwoven with interviews with experts and key actors in the controversy, participant observation of relevant events, as well as a document analysis (Asdal & Reinertsen, 2022) of the more than a thousand leaked documents, in a kind of 'digitally connective fieldwork' inspired by connective and multi-sited ethnographies (Hine, 2015; Marcus, 1995). In this way, I think of my project as *in-between* (Marres, 2017) or in the continuity of the old digitised methods and the new natively digital methods (Rogers, 2013, 2019). All this in order to develop a "cross-media analysis" (Venturini & Munk, 2022; Venturini & Rogers, 2019) of the connection between Meta's platforms and societal harms across four corpus of empirical materials from the disclosures, news media, corporate media, and social media, selected for providing ample material to analyse the multiple actors, issues and framings at play during the controversy.

Results

In this work-in-progress, I present some of my preliminary findings by examining how Haugen's disclosures were made public and which actors, issues and framings - and not others - gained greater prominence during the controversy to make or unmake connections between Meta platforms and societal harms. In the news and social media corpus I found a certain personification of the controversy, or a prominence of the whistleblower's figure and related events of the controversy over the discussions of the actual internal documents. Beyond this, it was promoted what I call a 'strategic causalism' in how the leaks were re-framed by actors in journalism, politics, and activism to solidify the connection between Meta's platform and specific harms. Moreover, in order to charge the disclosures with the capacity to solidify platform critique, these actors promoted causal connections between platforms and societal

harms - primarily the connection between Instagram and teen mental health harms- but always in a strategic way to achieve more 'feasible' platform regulations. And for that goal, "more scrutiny, the better", as one journalist told me.

In contrast, Meta's corporate response constantly sought to reframe the disclosures by emphasising nuances, giving more context, and explaining to us what their internal documents 'really say' (Raychoudhury, 2021). The research on platforms and mental health issues would be continuously framed as "relatively nascent and evolving" (Newton, 2021) to suggest that such a connection is not yet established or settled in resonance with Big Tobacco's discursive strategies (Proctor, 2006). Drawing on ignorance studies (Gross & McGoey, 2022), the Meta response promoted what I conceptualise as "strategic ambiguity." Instead of closing the controversy, Meta's executives would precisely try to state that the causal link between platforms and harms would still be an open controversy. But rather than denying negative effects, representatives would promote a utilitarian narrative that the negative effects would be outweighed by the positive effects of social media platforms in terms of social capital or connectedness. This appeal to nuances of platform harm research by Meta spokespersons can be understood as a dramatic technique to disperse responsibilities to others (e.g. users, malicious actors, politicians).

Finally, the re-framings across these media settings focused on a small number of issues that stand out strategically as those most likely to advance platform accountability regulations but nevertheless reinforce a U.S. and European centrism in the public trial about platform harms, that does not allow for more plural forms of platform accountability from and for the Global South. By moving from platforms as stages to putting platforms on stage, the personification and selection of certain issues over others restricted the possibility of discussing platform power and the increasing social engineering of sociality. In analysing the case of the Facebook Files/Papers, we are not only faced with the typical denunciation of wrongdoings by private companies, the spread of a increasingly critical tech journalism, or the difficulties of exposing such wrongdoings in the absence of platform accountability, but that the very idea of accountability must be reconsidered taking into account the asymmetries involved in the study and mediatisation of platform harms and who is qualified to give accounts of them.

References

Asdal, K., & Reinertsen, H. (2022). *Doing document analysis: A practice-oriented method*. SAGE Publications Ltd ; SAGE Publications Inc.

Boltanski, L. (2011). *On critique: A sociology of emancipation* (English ed). Polity.

Cooper, A. F., Moss, E., Laufer, B., & Nissenbaum, H. (2022). Accountability in an Algorithmic Society: Relationality, Responsibility, and Robustness in Machine Learning. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 864–876. https://doi.org/10.1145/3531146.3533150

Foroohar, R. (2018, December 16). Year in a Word: Techlash. *Financial Times*.

Goffman, E. (1986). *Frame analysis: An essay on the organization of experience* (Northeastern University Press ed). Northeastern University Press.

Gross, M., & McGoey, L. (2022). *Routledge International Handbook of Ignorance Studies: Second Edition* (2nd ed.). Routledge. https://doi.org/10.4324/9781003100607

Hendrix, J. (2021, October 25). *The Facebook Papers: An Anthology*. Tech Policy Press. https://techpolicy.press/the-facebook-papers-an-anthology/

Hine, C. (2015). *Ethnography for the Internet: Embedded, embodied and everyday*. Bloomsbury Academic, An imprint of Bloomsbury Publishing Plc.

Horwitz, J. (2021, October 1). The Facebook Files. *Wall Street Journal*. https://www.wsj.com/articles/the-facebook-files-11631713039

Jasanoff, S. (2019). Controversy Studies. In G. Ritzer (Ed.), *The Blackwell Encyclopedia of Sociology* (1st ed., pp. 1–5). Wiley. https://doi.org/10.1002/9781405165518.wbeosc130.pub2

Khan, L. M. (2018). Sources of Tech Platform Power. *GEORGETOWN LAW TECHNOLOGY REVIEW*, *325*.

Marcus, G. E. (1995). Ethnography in/of the World System: The Emergence of Multi-Sited Ethnography. *Annual Review of Anthropology*, *24*, 95–117. https://doi.org/10.1146/annurev.an.24.100195.000523

Marres, N. (2015). Why Map Issues? On Controversy Analysis as a Digital Method. *Science, Technology, & Human Values*, *40*(5), 655–686. https://doi.org/10.1177/0162243915574602

Marres, N. (2017). *Digital sociology: The reinvention of social research*. Polity.

Marres, N. (2021). No issues without media. The changing politics of public controversy in digital societies. In J. Swati & J. Wasko (Eds.), *Media: A Transdisciplinary Inquiry*. Intellekt Books.

Nelkin, D. (Ed.). (1992). *Controversy: Politics of technical decisions* (3rd ed). Sage.

Newton, K. (2021, September 14). *Using research to improve your experience*. Instagram. https://about.instagram.com/blog/announcements/using-research-to-improve-your-experience

Proctor, R. N. (2006). 'Everyone knew but no one had proof': Tobacco industry use of medical history expertise in US courts, 1990-2002. *Tobacco Control*, *15*(suppl_4), iv117–iv125. https://doi.org/10.1136/tc.2004.009928

Raychoudhury, P. (2021, September 26). What Our Research Really Says About Teen Well-Being and Instagram. *Facebook Newsroom*. https://about.fb.com/news/2021/09/research-teen-well-being-and-instagram/

Rogers, R. (2013). *Digital methods*. The MIT Press.

Rogers, R. (2019). *Doing digital methods* (1st edition). SAGE Publications.

Suarez Estrada, M., Juarez, Y., & Piña-García, C. A. (2022). Toxic Social Media: Affective Polarization After Feminist Protests. *Social Media + Society*, *8*(2), 205630512210983. https://doi.org/10.1177/20563051221098343

Van Dijck, J., Poell, T., & Waal, M., de. (2018). *The Platform Society. Public Values in a Connective World.* Oxford University Press.

Venturini, T., & Munk, A. K. (2022). *Controversy mapping: A field guide*. Polity.

Venturini, T., & Rogers, R. (2019). "API-Based Research" or How can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach. *Digital Journalism*, *7*(4), 532–540. https://doi.org/10.1080/21670811.2019.1591927