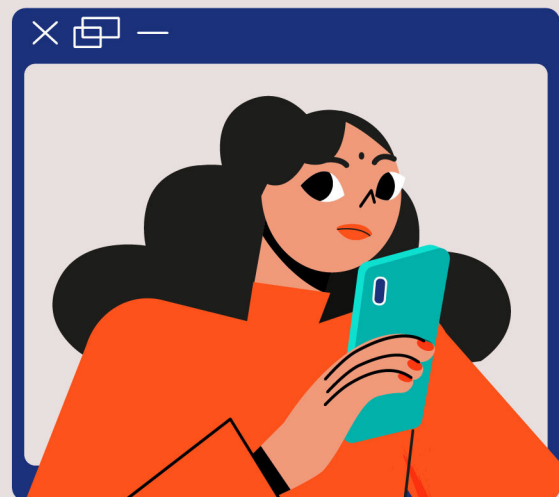
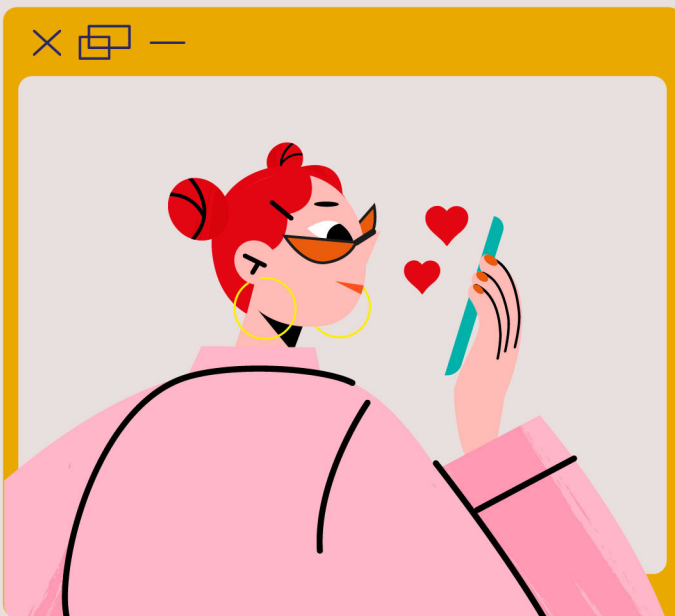


The Hidden Codes that Shape Our Expression:



**Understanding How Social Media
Algorithms Obstruct Feminist
Expression and How Malaysian
Women Navigate the Challenges**



Written By
Serene Lim

Edited By
Angela M. Kuga Thas

Copyright © 2022 KRYSS Network PLT
Creative Commons Licence: Attribution 4.0 International (CC BY 4.0)

ISBN Number: 978-967-17792-5-5

Published in 2022 by:

KRYSS Network
107A-A, Block A, Ground Floor,
Jalan SS2/72, 47300 Petaling Jaya,
Selangor.

Email: info@kryss.network
Website: www.kryss.network

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

This work forms part of the APC “Feminist Internet Research Network” project, supported by the International Development Research Centre, Ottawa, Canada. The views expressed herein do not necessarily represent those of IDRC or its Board of Governors.

Written by: Serene Lim
Edited by: Angela M. Kuga Thas
Cover and Layout by: Chan Qian Hui



e ISBN 978-967-17792-5-5





TABLE OF CONTENTS

INTRODUCTION	06
LITERATURE REVIEW	09
An Extended Understanding of Freedom of Opinion and Expression	01
Biases in Algorithmic Decision Making	01
Redesigning Our Expression For the Machine	01
Concluding Remarks	01
METHODS	14
DATA ANALYSIS	15
Shaping Public Discourse Through Algorithms	15
Resilience Against Hate-Inducing Algorithms	20
Riding The Hate	21
Resistance By Conforming	23
Denying Visibility To Hateful/Harmful Speech	23
Going Private	24
Audience Curation	25
Block, Mute and Report	26
CONCLUSION	28



INTRODUCTION

Women's¹ visibility and expression on social media are often burdened by the risks of hate and harassment. The potential backlash and violence online have spurred women in all diversity to practise self-policing and censorship and to base their expression on the perceived reactions of their audiences.² Several women in KRYSS Network's earlier research have also expressed that the fear of disparagement and vitriol had led them to modify the way they expressed and spoke in digital spaces, knowing that they have very little to no control over their narratives once they become a target of online gender-based violence.³

While online gender-based violence is rooted in gender-based discrimination that takes place in every facet of society, this research seeks to understand how such form of violence might be facilitated in particular ways by the algorithm and design of social media platforms. The design of social media is not neutral but is planned, prototyped and developed to invite and shape participation toward particular ends, including what is not permitted and the policing of objectionable content and behaviour.⁴ However, what is not permitted and the objectionable content and behaviour that is policed, are not necessarily to promote and protect human rights.

Recent debates around freedom of opinion and expression on social media have expanded beyond content moderation to the way algorithms have come to interfere with the flow of information, amplifying some and suppressing others. All of our expression on social media is subjected to an algorithm that amplifies or suppresses its circulation to maximise data extraction vis-à-vis profit for social media platform owners. This ultimately has a direct

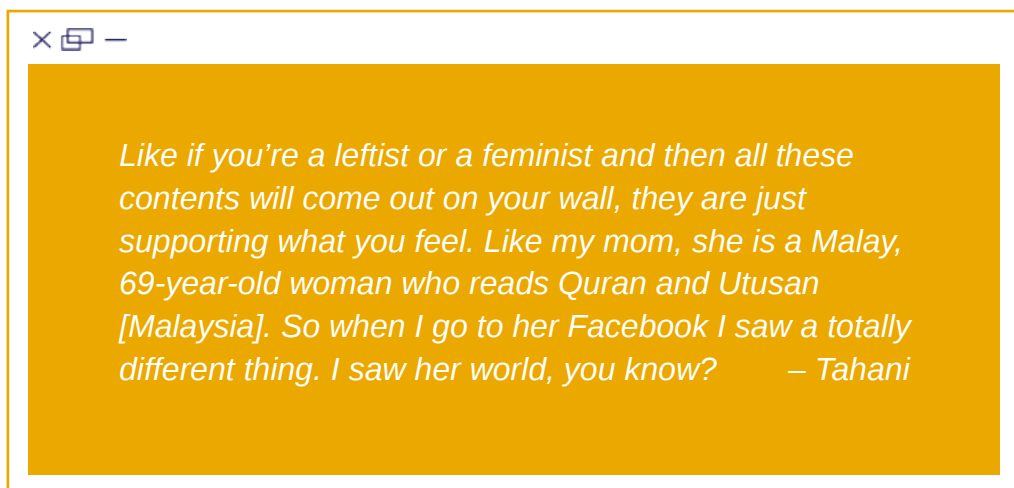
1 The term "women" is used to include cisgender women, transgender women and female-presenting who identify as non-binary.

2 Duffy, B. E., & Hund, E. (2019). Gendered Visibility on Social Media: Navigating Instagram's Authenticity Bind. *International Journal of Communication*, 13(2019), 4983-5002. <https://ijoc.org/index.php/ijoc/article/view/11729>

3 Lim, S. (2021). *Power X Expression X Violence: A Research on Women's Freedom of Expression on Social Media in Malaysia*. KRYSS Network. <https://firm.genderit.org/research/power-x-expression-x-violence-womens-freedom-expression-social-media-malaysia>

4 Munn, L. (2020). Angry by design: toxic communication and technical architectures. *Humanities & Social Sciences Communications*, 7. <https://doi.org/10.1057/s41599-020-00550-7>

influence on who gets to be heard and who gets to hear what speech. In this sense, our freedom of opinion and expression is not free *if we can speak but not be heard*.⁵ The algorithmic power in determining “who should be heard” and “what should you read” is not just technical architecture but is imbued with capitalistic and patriarchal logic that jointly reinforces oppression against women and gender non-conforming persons.



Underlying Tahani’s sharing is the appropriation and treatment of public discourse as mere content and data by social media companies with the aim to increase engagement and drive advertising revenues. Our expression and information online are now a means to commercial ends. The result is the inevitable preference for content that is populist and sensational, and reinforces the values of the status quo including racist and misogynist beliefs.

5 Riemer, K., & Peter, S. (2021). Algorithmic audiencing: Why we need to rethink free speech on social media. *Journal of Information Technology*, 36(4), 409-426. <https://doi.org/10.1177/02683962211013358>



LITERATURE REVIEW

With more discourse and expression of ideas and thoughts taking place on social media, social media companies now wield immense power in organising, influencing and controlling how freedom of opinion and expression can be exercised, and how each expression is presented, or not, to each of us. Other than the acts of censorship and content moderation, the algorithm presents a novel form of control through which our expressions are commodified and distributed algorithmically, primarily underpinned by the economic logics of monetisation but catering too to how the manifestation of such opportunities is very much rooted in patriarchal cultures and unequal gender power dynamics. This literature review will discuss how the platforms' algorithm interferes with our ability to equally exercise and access the right to freedom of opinion and expression by determining what expressions get priority on their platforms, often severely affecting women and gender non-conforming persons disproportionately.

An Extended Understanding of Freedom of Opinion and Expression

The debate around the role of social media in shaping freedom of opinion and expression (referred to as “freedom of expression”

hereafter) often revolves around content moderation and censorship. In KRYSS Network’s earlier research, it was also raised that social media platforms often engage with content moderation models that are deeply averse to freedom of expression. Harassment and violence that are gender-based are often framed as “user-generated content” under their terms of service.⁶ Other than the power to police and govern what kind of expression is allowed or disallowed, these social media companies are actively interpreting our expression as data and content, deciding who gets to hear and read our content in opaque and non-transparent ways. From “what can be said” to “what will be heard and by whom”, there is a dire need to extend our debate around freedom of expression to include algorithmic interference in the exchange and flow of ideas and its impact on public discourse.

As a starting point, it is important to emphasise that freedom of expression is not merely about the ability to speak, but also the right to be heard.⁷ The powerful are free to speak and they are given the free rein to silence the vulnerable by reinforcing false stereotypes that undermine their credibility or by intimidating them. For those who belong to stigmatised and marginalised groups and communities, when they speak, their voices are often ignored, or the audiences

6 Lim, S. (2021). Op. cit.

7 Riemer, K., & Peter, S. (2021). Op. cit.

8 Bell, M. C. (2020). John Stuart Mill’s Harm Principle and Free Speech: Expanding the Notion of Harm. *Utilitas*, 33(2), 162-179. <https://doi.org/10.1017/S0953820820000229>

9 Riemer, K., & Peter, S. (2021). Op. cit.

may discount their stories or try to silence them.⁸ Freedom of expression is also about discourse,⁹ and this freedom plays a public role by giving equal opportunity to everyone to participate in discussions, thus contributing to the accumulation of collective knowledge and social progress.¹⁰ Inevitably, any form of interference with the free exchange and debate of ideas and different viewpoints is harmful to the intellectual and collective development of societies.

“Algorithm audiencing”, coined by Kai Riemer and Sandra Peter, refers to the automatic and ad hoc configuration of audiences for speech through algorithmic content distribution, as a by-product of profit maximisation.¹¹ Using Facebook (now Meta) as a case study, the authors further elaborate that the algorithm takes our speech out of its context and determines circulations based on criteria and metrics that will benefit Facebook’s bottom line. Algorithmic decision-making is designed to measure what is trending and what might interest us, suggesting or pushing content, amplifying some expression while suppressing others. In this regard, social media companies hold immense power as they can, will and do, directly interfere with freedom of expression by creating a false perception of free and equitable access to audiences through amplification or suppression of speech, and by doing so, determine the size and characteristics of audiences that get to see a particular message.¹² Together, they not only manipulate the flow of information and

discourse on our social media page, but can and do contribute to reinforcing existing biases, prejudices, stigmatisation, gender-based violence and more.

Biases in Algorithmic Decision Making

Algorithms are meaningless machines until they are paired with a database that sets encoded procedures or instructions to transform the data inputted to the machine into the desired output, based on specific calculations.¹³ Despite the promises of objectivity, comprehensiveness and reduction of human biases, the process of collecting and analysing large amounts of data includes human choices in either excluding information from a database or including and then managing them in particular ways.¹⁴ Algorithmic biases are not always done intentionally but can be a result of unconscious biases by the designer who often default their imagined users to those based on their values, lived realities and singular world view. This means most of the existing tools are designed to benefit members from the dominant group, i.e. cis, male, white, heterosexual, able-bodied, literate, with good internet access, etc.¹⁵

Various studies have shown that gender bias and discrimination in the design of digital technologies—its structures and features—are pervasive and have direct negative impacts on women and gender non-conforming persons’ equal access

10 Bell, M. C. (2020). Op. cit.

11 Riemer, K., & Peter, S. (2021). Op. cit.

12 Ibid.

13 Gillespie, T. (2014). *The relevance of algorithms*. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media Technologies: Essays on Communication, Materiality, and Society*. MIT Press.

14 Ibid.

15 Constanza-Chock, S. (2021). *Design Justice: Community-Led Practices to Build the Worlds We Need*. MIT Press

to opportunity and the defence of their human rights.¹⁶ Social media algorithms can reinforce and amplify existing harmful gender stereotypes and inequalities. For instance, an audit by researchers at the University of Southern California (USC) showed that Facebook's advertisement delivery system shows different job ads to women and men even though the jobs required the same qualifications—suggesting the jobs are selectively shown to people of different gender based on the current demographic distribution of these jobs.¹⁷

Data extraction and the encoding procedure in these algorithms neither exist in a vacuum nor do they exist outside of the social world. Decisions about what to include and what to ignore at the design level, what to pay attention to and what to disregard during data collection and analysis are always guided by everyday power relations situated in gender, ethnicity, religion, age, socio-economic status, etc.¹⁸ Gender stereotypes and discrimination can also operate outside of our conscious awareness.¹⁹ What is often taken to be “natural” and “neutral” by technologists and mathematicians can be discriminatory beliefs and stereotypes that have been normalised in our society, and do not reflect the lived realities of women, girls and gender non-conforming persons. As shown by the research on the Facebook

ads, without considering one's merits or qualifications, the algorithms presume women are more likely to engage with job opportunities that have more women in the current demographic distribution (i.e. grocery delivery, jewellery, domestic work, etc.), which often differ based on historical and existing gender-based discrimination.²⁰

Redesigning Our Expression For the Machine

The other way in which algorithms influence our freedom of expression is how we have oriented ourselves towards them to ensure our contents and expressions are algorithmically recognised and boosted, as we seek to monetise or achieve a broader reach for our content.²¹ Countless resources have emerged to guide users on how to craft their message effectively so that these posts are able to reach more audiences on social media. It is not about the value or merit of one's expression but how well a certain message, as content, can secure higher engagement potential with a wider audience. These often include highly emotive and often harmful and inflammatory contents.²²

For feminist content creators, there exists a tension between resistance and compliance with the status quo as they structurally and technically adjust, conceal and amplify an

16 Suzor, N., Dragiewicz, M., Harris, B. A., Gillett, R., Burgess, J., & Van Geelen, T. (2018). Human Rights by Design: The Responsibilities of Social Media Platforms to Address Gender-Based Violence Online. *Policy & Internet*, 11(1), 84-103. <https://doi.org/10.1002/poi3.185>

17 Hao, K. (2021, 9 April). *Facebook's ad algorithms are still excluding women from seeing jobs*. MIT Technology Review. <https://www.technologyreview.com/2021/04/09/1022217/facebook-ad-algorithm-sex-discrimination>

18 Kovacs, A., & Ranganathan, N. (2019). *Data sovereignty, of whom? Limits and suitability of sovereignty frameworks for data in India*. Internet Democracy Project. <https://internetdemocracy.in/reports/data-sovereignty-of-whom>

19 Noll, N. (2020, 17 February). *Gender Equality ≠ Gender Neutrality: When a Paradox is Not So Paradoxical, After All*. *Gender Sci Lab*. <https://www.genderscilab.org/blog/gender-equality-does-not-equal-gender-neutrality>

20 Hao, K. (2021, 9 April). Op. cit

21 Gillespie, T. (2014). Op. cit.

22 Roose, K. (2021, 9 January). *In Pulling Trump's Megaphone, Twitter Shows Where Power Now Lies*. *The New York Times*. <https://www.nytimes.com/2021/01/09/technology/trump-twitter-ban.html>

aspect of themselves to fit into the social media's socio-techno order.²³ The prevailing heteronormative norm on social media often means that algorithms will prioritise contents that project relatable femininity through neoliberalism and a patriarchal lens. In some ways, feminists have had to comply with practices that are against their political convictions or hold back on their political stance because these are insufficiently popular for social media, especially when they want to ensure the message goes out without filtering or obstruction. If not, they risk being unheard and remain in the margins of visibility or in echo chambers.²⁴

KRYSS Network's earlier research also illustrates how women and gender non-conforming persons are rendered much more vulnerable by the algorithms created by social media platforms. The visibility accorded to women by algorithms in terms of engagements and followers is a double-edged sword. The internet and social media in general provide women better access to public participation but it also means being subjected to intensified scrutiny publicly, and women are expected to express themselves in accordance to gender norms, failure of which will lead to vitriol attacks. In these situations, women engage in strategic acts of revelation and concealment of self, a delicate form of self-expression and self-protection that have come to anticipate networked hate.²⁵

Concluding Remarks

Algorithmic interference in freedom of expression is an emerging and important area of research. It can drive polarisation, reinforce existing disparities and discriminatory practices, and has a larger societal concern over manipulations of the distribution of information. Despite the power asymmetries between big corporations and communities at risk, along with the obscurity of algorithmic decision-making, it is encouraging that the participants of the earlier research had begun to identify and develop strategies and tactics to subvert these algorithms. This paper intends to better understand the barriers and biases resulting from algorithms in women's access to freedom of opinion and expression, and to examine women's resiliency and how they navigate these algorithms that are inherently limiting to create the much-needed space for women and gender non-conforming persons to speak out, to be heard, and to, in effect, occupy digital spaces.

²⁴ Ibid

²⁵ Lim, S. (2021). Op. cit.



METHODS

The origin of this paper came from our earlier research, “Power X Expression X Violence: A Research on Women’s Freedom of Expression on Social Media in Malaysia,” in which we analysed freedom of opinion and expression as a discourse of power—how it reinforces gender inequalities through unequal access to freedom of expression. Through our interviews, when the women and gender non-conforming persons talked about the barriers to their freedom of expression, it soon became clear that the social media algorithms rendered them much more vulnerable. Hate speech and violence are profitable when the focus is only on increasing viral content. The design of these algorithms very much suggest a male/patriarchal gaze on what is desirable content, and it is encouraging that the participants of the research had begun to identify and develop strategies and tactics to subvert these algorithms.

Data from this paper came from a re-examination of interview transcripts of 23 women and gender non-conforming persons who were targeted with online gender-based violence and five individuals who had participated in perpetrating online harassment. These interviews were conducted from April to November 2019 for an earlier research on the inherent inequalities in women’s access to freedom of opinion and expression and how their exercise of this freedom invites online gender-based violence. The richness of these interviews means there is much to understand about women and gender non-conforming persons’ lived experiences and resilience in digital spaces and hence the conception of a second research paper based on the stories obtained from these interviews. In addition, the author also included four case studies documented and developed by KRYSS Network as part of the organisation’s advocacy to eliminate online gender-based violence.



DATA ANALYSIS

This part of the paper is divided into two main sections. The first section examines how the algorithm, compounded by inherent structural gender-based discrimination, is shaping and interfering with women and gender non-conforming persons' equal access to freedom of opinion and expression, which in turn enables the amplification of online gender-based violence and hate speech. The second part looks at how the algorithms are adopted by women and gender non-conforming persons' in their expression of self digitally, imbuing them into their resistance against norms and discriminatory practices.

01

Shaping Public Discourse Through Algorithms

The algorithms that power the flow of information and exchange of discourse on social media are designed to capture people's attention by constantly tracking and predicting our behaviour, and in return, they feed us contents that are most likely to interest us—to encourage us to “like” or share or further interact with the content. In an era of the attention economy, human attention is deemed a scarce resource and a form of currency—the more followers users have on their social media accounts, the better they can benefit from them economically, socially or politically. Yet,

users are afforded little to no control over the distribution of their content, the type of audiences they are reaching, or the type of attention they get. These algorithms and digital infrastructures weave themselves into our lives and influence the way we express and what we see. Yet, the design and decisions on how they function is unknowable to us.²⁶

When asked what she hoped to change about social media, Nadia shared that she wished to have the ability to customise who can see her tweets or contents on Twitter, and to control the kind of reactions she received to her tweet. When Mei posted a mundane video of her dancing in a cartoon onesie without wearing a bra, on TikTok, it went viral and she woke up to a million viewership for that particular video, accompanied by thousands of sexually offensive and slut-shaming comments, all focusing on her breasts. Overnight, she received up to a million interactions and thousands of followers who were only interested in her as a sexual object, qualifying her for the local TikTok creator programme. At the same time, the interactions and comments for her subsequent videos (especially for those where she spoke about feminism and social justice issues) remain low and sometimes as low as 500 views for an account with 71,000 followers.²⁷ In an interview with her, she expressed how the said video has messed

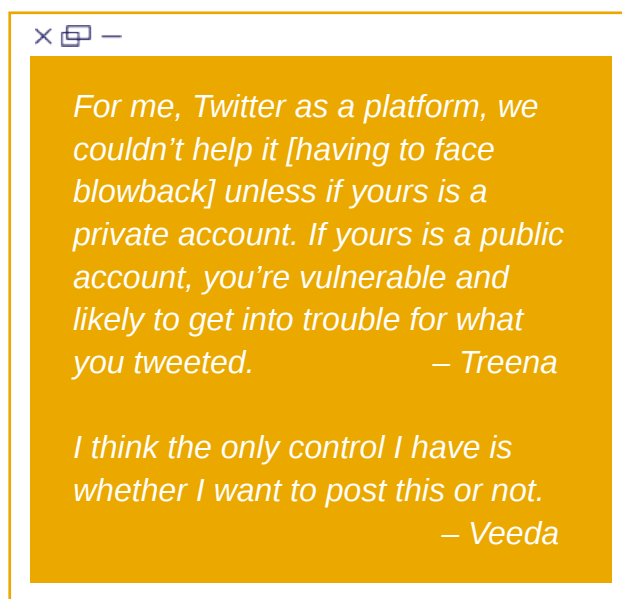
²⁶ Zuboff, S. (2019). *The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power*. Profile Books.

²⁷ The numbers are recorded as of 12 September 2021.

up her TikTok account whereby now she has more followers who are intent on sexualising her rather than to appreciate or understand the original message of her contents. She expresses that she does not know how else she can reclaim and redirect her account to her intended audience.

Sexual objectification of the female body is an age-old tale and has found a way to manifest online. This form of sexual oppression is further compounded when Mei's body is reduced to a data point by the algorithm that decided the video should be made visible for the male/patriarchal gaze, leaving her without any agency as to how her body should be treated or viewed because of the algorithm.

Given the power imbalance, most research participants shared the consensus that they do not have much control over the distribution of their content. Treena's public tweets on her feminist politics often spurred blowback and it gets exhausting for her to defend and engage.



The visibility of the #MeToo movement contributed to the high number of engagements Nadia received when she tweeted her own #MeToo sexual harassment stories as a woman with a disability. Before this, her usual tweets on her disability-related stories received very little traffic or engagement. She first received words of encouragement on her #MeToo tweet but was soon bombarded with hateful and sexist comments from strangers. The key to understanding Nadia's experience is the question of what is prioritised by the algorithm, what got pushed to the top of the timeline to capture Twitter users' utmost attention; what contents are deemphasised and buried to the bottom of the page or are being excluded altogether, and what are the factors considered by the algorithm when curating content for the users. Most social media algorithms are driven by engagement. For Facebook, the factors include who is posting, the frequency of posts and the average time spent on the content.²⁸ Much like Facebook, engagement rate, i.e. the number of retweets, clicks or impressions a tweet has received, types of media included in one's tweet and who is posting it, all play a role in Twitter's timeline ranking signals.²⁹

The underlying logic of algorithms to amplify "engaging" content means that any attacks or harassment from aggressors, regardless of its potential harm, will send a positive signal to the algorithms. Once it has been ranked up by the algorithms for amplification, it is usually too late for anyone to try to remove the initial content or contain the blowback, personal attacks, and undesired sexualisation. Amy's tweet received users'

27 The numbers are recorded as of 12 September 2021.

28 Munn, L. (2020). Op. cit.

29 Nemeth, C. (2020, 11 February). What to know about how the Twitter algorithm works. *Sprout Social*. <https://sproutsocial.com/insights/twitter-algorithm>

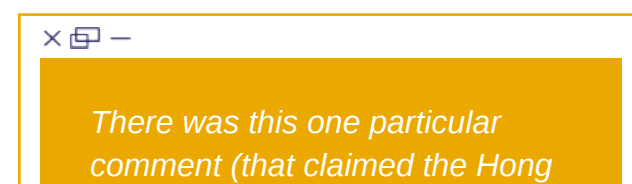
wrath after she called out the mufti (Islamic jurist) of Perak's behaviour of bullying young Malay Muslim girls for mourning over a K-pop artist. At the time of posting, she had no reason to believe that her tweet would gain such traction as she only had a low number of followers. When she saw a known anti-feminist troll engage with her tweet, she knew it would be pointless to report the harassment or attempt to defend herself as his high number of followers would have drawn in more attacks against her. Even if she did manage to have the one tweet removed as violating content norms, there were thousands of replies and quote tweets linked to the violating tweet and she simply had no energy to report each tweet one by one. More so, people had already taken screenshots of her tweet, posting it on their own page and their respective followers were engaging and replying to the tweet, without tagging or reference to her original tweet.

The prioritisation of engagement by social media algorithms stems from the underlying capitalistic logic of opportunity to capture users' data in order to be better able to predict their behaviours, but still heavily premised on the patriarchal and heteronormative understanding of what these behaviours could/should be. Social media algorithms present contents that are more likely to keep them more engaged and for longer, and this in turn enables data extraction, i.e. age, gender, location, political ideology, preference, etc. These behavioural data are then traded into prediction products, with targeted advertising being one of the dominant business models in the earlier days.³⁰ Such logic stands in

direct conflict with the need of constructive, open, diverse and substantive dialogues for a thriving democratic society. The ability to rank and control the distribution and visibility of a particular content is a powerful one and it is a power that is hidden from all of us. The interference of algorithms by amplifying certain expressions to the audience by virtue of their ability to engage vis-à-vis marketability for behaviour data, will invariably distort the free exchange of opinions and expose women to more violence and hate.

Online gender-based violence, while having its roots in existing structural gender inequalities, in part is a consequence of the algorithm's underlying logic to produce high engagement. Misinformation, and outrageous, polarising and sensational information, typically produce high engagement and dominate the top of social media timelines.³¹ In the last few years, the drawbacks of such designs have been publicly expressed by designers from social media platforms—how the design of Facebook is “ripping apart the social fabric of how society works” with the “short-term, dopamine-driven feedback loops” that discourage civil discourse but reward violence, misinformation and untruths.³²

Albert, in describing how he felt when he read a post on the Hong Kong Umbrella Movement on Facebook, shared:



30 Zuboff, S. (2019). Op. cit.

31 Levy, S. (2020). *Facebook: The Inside Story*. Blue Rider Press.

32 Meyer, D. (12 December 2017). Facebook is 'ripping apart' society, former executive warns. *Fortune*. <https://fortune.com/2017/12/12/chamath-palihapitiya-facebook-society/>

Kong police officers were violent), it was very silly and I feel I have to say something (about how they are injuring public servants), I don't know why, I just decided to post something. [...] I think I wasn't being myself, it was done out of anger.

In a separate incident, Albert, being of Chinese-Malaysian descent in a Malay-Muslim majority country, found strong resonance with a documentary of African-American boys who were wrongly convicted and jailed for crimes they didn't commit. He was outraged and turned to Twitter to find out if others had mobbed the judge who wrongly convicted the boys. He retweeted some of the tweets attacking and condemning the judge, who by then had already lost her job and deactivated her Twitter account. Though he was not actively saying anything, he felt like he was part of the mob.

✕ 🗉 –

At the end of the day, you are still attacking someone [...] and she wasn't able to defend herself. I think there are 10,000 handles against one person. I feel I was part of the mob, honestly. And, it's very stupid also because you know, I'm this Chinese guy from Malaysia living a thousand miles away from her, she doesn't know, and I know about her from watching the documentary, reading from one side and I decided

to just jump in and support the other side.
– Albert

Moral outrage is an all-too-familiar and arguably the most viral emotion on social media. While the occurrence of the outrage mob on social media is not isolated from existing social and political issues, the algorithms create a process similar to positive reinforcement learning where outrage contents or activities are “rewarded” with more likes and reshares.³³ Algorithms can also incentivise outrage through the communication norms presented to users in their networks, where users subconsciously adjust their behaviour by following what the majority are doing. In Albert's case, his initial emotional reaction was validated when he saw that other users had started harassing and attacking the former judge.

Reflecting on the online harassment against her during the women's march in 2018, Nina now knows that hate and violence spread so quickly in a short amount of time. “So now I really realise how important it is to ask consent and also how to censor a few things that people might not want out there because of these kinds of things,” she shared.

Digital spaces are an extension of our physical world and the proliferation of gender-based moral outrage on social media is an inevitable result of sexist, misogynist, homophobic and transphobic societies. That being so, automated interventions alone, i.e. content moderations using artificial intelligence by social media companies will always be inherently limited as they will not be able to account for the nuances

33 Kalia, S. (2021, 16 August). How Retweets and Likes on Social Media 'Reward' Moral Outrage. *The Swaddle*. <https://theswaddle.com/how-retweets-and-likes-on-social-media-reward-moral-outrage>

of particular contexts, power dynamic imbalance and unequal access to freedom of expression.³⁴ Further facilitated by hate-inducing architectures, moral outrage is particularly contagious on social media, and the seamless sharing features on social media, just a click away, have allowed such content to rapidly proliferate across the network.³⁵ In an interview, Faiz, one of the aggressors, said it was easier for him to attack or curse someone on Twitter, something that he would not do in person.

x [icon] —

Online is easier because you can just disperse whenever you want. [...] If you have an argument with someone offline right, usually it's with someone you know. You don't go up to a stranger and say "Weh, I don't agree with atheism lah." So because you have the personal connection that you have, so it's really hard for you to have a conversation that might offend the person of interest. — Faiz

Faiz further shared his observation on how others from his network use religion and conservative narratives to gain attention and followers.

Sometimes if you have a lot of followers, you can do reviews and get paid. So, for me, it's really not surprising to see people tweeting religious or ideological stuff to actually gain attention, because attention is money. So why not right? — Faiz

Faiz's observation is also echoed by Kazim, who had admitted that he used to troll those who identified themselves as feminists or LGBTIQ persons on Twitter but has now gained a better understanding of gender identities and sexual orientation. At the time of the interview, he had about 10,000 followers on his Twitter and he stated that he was a lot more expressive back when he only had half of the followers. The reward system based on one's followers and influence provides very direct incentives to how one expresses or reacts to political events over time. In Kazim's case, it was difficult for him to speak up against homophobic narratives as he now risks being attacked and losing followers thereby potentially affecting his income. In this way, social media algorithms trap users into a way of being and expression; there is no room for expressing personal growth, improved understanding of issues—political, social, economic or technological; and the change that person has experienced in beliefs, values, attitude and perception. To that extent, social media algorithms can be said to work against improving just basic civic consciousness in societies.

02

Resilience Against Hate-Inducing Algorithms

Foregrounding the power imbalance presented by the gender-biased algorithms is strength and resiliency of women and gender non-conforming persons to cope with gender-based violence, and in some cases, to ride on the hate. The next part

34 Munn, L. (2020). Op. cit.

35 Ibid.

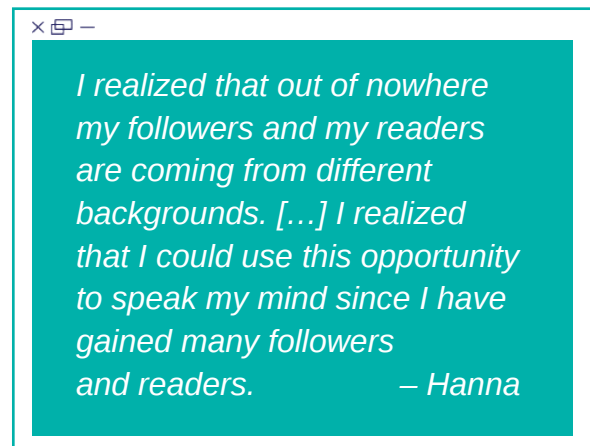
of the paper exemplifies different coping mechanisms employed by women and gender non-conforming persons to uphold and protect themselves through resilience, despite the structural inequalities that limit them. In our analysis, we identified six different ways in which women enact resiliency:

- **Riding The Hate**
- **Resistance By Conforming**
- **Denying Visibility To Hateful/Harmful Speech**
- **Going Private**
- **Audience Curation**
- **Block, Mute And Report.**

Riding The Hate

Ironically, women often found themselves gaining algorithmic visibility and followers after an incident of online gender-based violence. Katherine gained more than 1,000 followers on Twitter after an online mob targeted her. She would search her name on Twitter and found random people talking about her with most of them insulting her, claiming that she intentionally caused provocation to gain followers. This too is an interesting observation as those who attacked her online also knew that deliberate provocation can effectively increase a Twitter user's number of followers. Fortunately, the increase in visibility and followers had a liberating impact on Katherine's freedom of expression. She described herself as more outspoken now and feels encouraged to speak truth to power. Hanna had a similar

trajectory when she started making a Facebook post about polygamous marriage public and was subsequently harassed online. She described the event as a turning point.

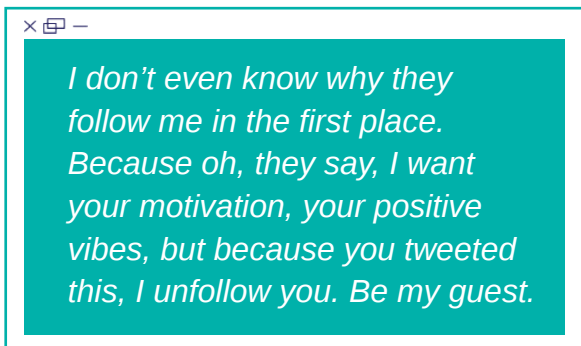


Instead of resisting the visibility afforded by the algorithms, Katherine and Hanna exhibited resiliency and were able to adapt to the incident of violence and ride on the system to further amplify their voices.

The upbeat framing of visibility and followers must be treated with caution. With Hanna, she felt she had the thorough knowledge of Islamic texts and a background in law to be better able to argue her points, and to have the confidence to hold steadfast to her views and position. For Katherine, it is her involvement with human rights organisation since her university years and law education that allow her to stand by her opinions. So other factors can and do play a role for these to see visibility as a progression in their activism. For others, it is tantamount to surveillance and increased exposure to hate and harassment. When Mia's photo featuring herself and her signboard at the 2020 Women's March first gained traction

on Twitter, even as she felt flattered by the compliments on her signboard, she was also afraid of the unwanted attention from strangers and the emotional burden that comes with troll-like and sexist responses. When Maimuna was attacked for standing up for LGBTQ rights, people from the queer community also accused her of bringing too much attention to the community, though she was the one who bore the brunt of the violence.

Sadia, who already had about 7,000 followers at the time of the attack, lost about 200 followers over two days when she pointed out the unfair treatment by a husband who took his time to eat and smoke before taking the baby from his wife so that she could eat her food. One of her followers even messaged and demanded she remove the tweet, or else she would unfollow her. Sadia recalls:

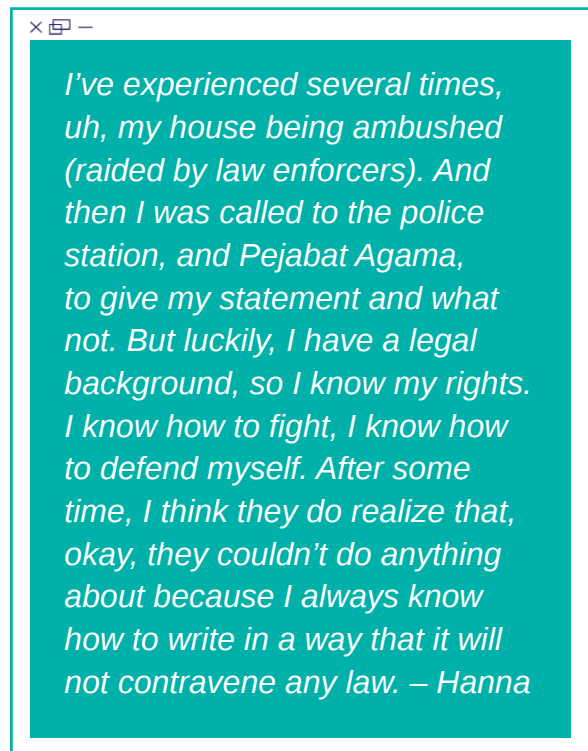
A screenshot of a tweet with a teal background and white text. The text reads: "I don't even know why they follow me in the first place. Because oh, they say, I want your motivation, your positive vibes, but because you tweeted this, I unfollow you. Be my guest." The tweet is enclosed in a thin teal border.

I don't even know why they follow me in the first place. Because oh, they say, I want your motivation, your positive vibes, but because you tweeted this, I unfollow you. Be my guest.

Most of her followers are Malay Muslims and that tweet was said to be an attack on the status and needs of Malay Muslim men. She is aware that her followers are likely to be uncomfortable or even opposed to the idea of gender equality, yet she felt she had to speak up about discrimination Malay Muslim women experience in their everyday lives, even if it meant losing more followers. For

reasons unknown to her, the harassment led to an unintended increase in her online business of selling religious items, though, at the same time, many were mobilising others to boycott her business.

The ability of Katherine and Hanna to resist and push back against gender-based violence should also be contextualised against their identity as cis heterosexual women, and their ability to access the justice system and awareness of their rights under the national laws. Resilience is a dynamic process dependent on individual traits or ability to cope with trauma, and a range of ecological factors including family, school, peers, community responsibility and social justice, through which survivors recover or move forward from adversity and violence.³⁶ In other words, resilience is a political and complex construct involving not only the individual but also multiple interacting systems, i.e. social, cultural, judiciary, and economy.

A screenshot of a tweet with a teal background and white text. The text reads: "I've experienced several times, uh, my house being ambushed (raided by law enforcers). And then I was called to the police station, and Pejabat Agama, to give my statement and what not. But luckily, I have a legal background, so I know my rights. I know how to fight, I know how to defend myself. After some time, I think they do realize that, okay, they couldn't do anything about because I always know how to write in a way that it will not contravene any law. – Hanna" The tweet is enclosed in a thin teal border.

I've experienced several times, uh, my house being ambushed (raided by law enforcers). And then I was called to the police station, and Pejabat Agama, to give my statement and what not. But luckily, I have a legal background, so I know my rights. I know how to fight, I know how to defend myself. After some time, I think they do realize that, okay, they couldn't do anything about because I always know how to write in a way that it will not contravene any law. – Hanna

36 Tsirigotis, K., & Luczak, J. (2018). Resilience in Women who Experience Domestic Violence. *Psychiatric Quarterly*, 89(1), 201-211. <https://doi.org/10.1007/s11126-017-9529-4>

Similarly, Katherine, who was studying to become a lawyer, has lodged a police report against the violent and hate speech against her. Even though her police report was not taken seriously by law enforcers, the ability to have access and make the police report was, to some extent, empowering for her. It is important to note that Katherine was the only one among the 23 women interviewed who lodged a police report on her own.

No doubt that there may exist other factors, both internal and ecological, that affect women's resiliency, where further study will be needed. That aside, the stories from Katherine and Hanna further illustrate the availability and accessibility of laws as an important condition towards building women's resiliency and sense of agency, which has very direct impact on their ability to fully access the opportunities and resources accorded by the internet.

Resistance By Conforming

While it is important to account for the power asymmetries of data collection and algorithmic content distribution, we should not imply that women are passive users and powerless against these tools. Social media sites offer spaces for all of us to present ourselves through profile building, the accounts we follow, the expression of our preferences or what interests us the most, etc. As a body-positivity advocate, Bonnie started creating more sex education content using her own photos after noticing that posts with her face received a higher

engagement compared to those without. The visibility afforded by social media allowed her to build her own community and connect with those who resonate with her content, and in return, she felt her freedom of expression is encouraged and celebrated.

Bonnie believes that Instagram is not the place to show negativity or frustration and finds it hard to talk about her support for LGBTIQ rights openly. Similarly, for Nadia, the expression of self on social media is curated and designed in a manner to avoid harassment and negative reactions among audiences. Rather than seeing this as a form of self-censorship or not "keeping it real", both Bonnie and Nadia see it as a form of strategy where conforming to the norms and rules of social media helps in gaining visibility and followers, and expressing their political opinions strategically. For Bonnie, it is a balancing act between challenging patriarchal norms and heteronormativity and the desire to get visibility and recognition in the form of likes and comments. For instance, instead of publicly declaring her support for LGBTIQ rights, she talks about diverse sexual orientations and how it is different for everyone.

Social media and their algorithms are yet another terrain where conformity and resistance to the status quo manifest as women strategically adopt, conceal, and amplify an aspect of themselves and their expression. This form of resistance, like Bonnie's, is subtle and unobstructed, and equally important in effecting mindset change.³⁷

37 Ybema, S., & Horvers, M. (2017). Resistance Through Compliance: The Strategic and Subversive Potential of Frontstage and Backstage Resistance. *Organization Studies*, 38(9), 1233-1251.

Denying Visibility to Hateful/ Harmful Speech

“Don’t feed the trolls” – a reflexive piece of advice given to women facing a deluge of harassment and violence online. Most of the women agreed that it is futile to try to engage with trolls or to argue with them. Self-care was cited widely as the reason to not engage the trolls. Maimuna believes the trolls are not there for the dialogue but merely to disrupt, annoy and spam. More importantly, ignoring the trolls also means denying them any further visibility. Any engagement, regardless of their motivations and purposes, only trick the algorithms into pushing it into a broader array of users’ timelines and newsfeeds.³⁸

The number of followers is also one of the main factors Veeda and Treena took into consideration when deciding whether to respond to comments online. They would usually refrain from engaging with an account with a massive following and conservative and right-wing opinions.

If I engage with them, am I giving them a platform? Does that mean their 500 followers will become 510, which is 10 too many from that engagement.
– Veeda

It is not only the one account she has to deal with, but her responses would also be made visible to her followers, who are very likely to share similar bigoted viewpoints. Twitter’s algorithm also includes one’s number of

followers when ranking the priority of each tweet.³⁹ For Veeda, the decision to not engage trolls is also to not lend visibility to hateful and harmful tweets to people within her network.

The economic logic of social media algorithms is aptly summarised by one of the research participant’s observations on the attention economy. Lily believes the most effective step she can take is to not engage with the trolls and not be exploited by the attention economy.

Attention itself is an economy. A value. And with enough attention, you are automatically an opinion leader. [...] Unfortunately what is effective [to get followers] is sexist and religiously prescriptive, and racist content.
– Lily

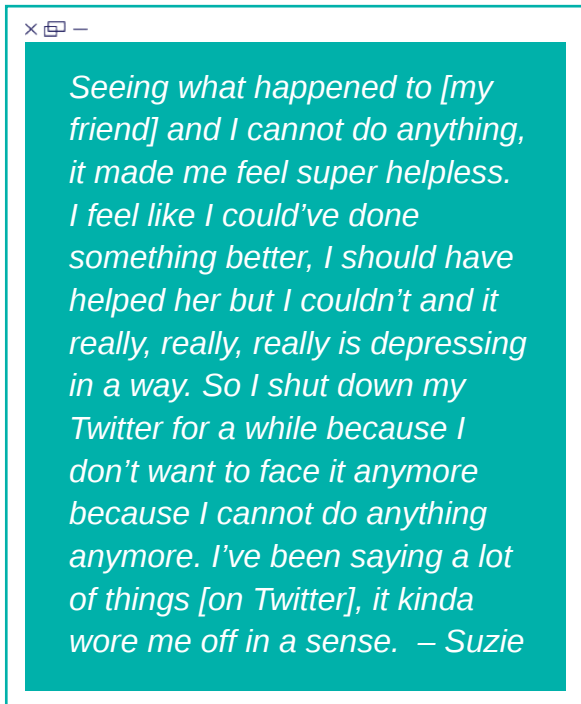
Going Private

Instead of feeding the algorithms by engaging with the trolls publicly, some women have chosen to privately reach out to those who are publicly harassed and attacked through WhatsApp or Telegram. Zara shared that she preferred to show support and solidarity offline or through a more private channel. She explains, “I’m not the one to wade into the Twitter wars, I can’t do that.”

³⁸ Center for Countering Digital Hate. (2019). *Don’t Feed the Trolls: How to deal with hate on social media.* https://docs.wixstatic.com/ugd/f4d9b9_ce178075e9654b719ec2b4815290f00f.pdf

³⁹ Nemeth, C. (2020, 11 February). Op. cit. f

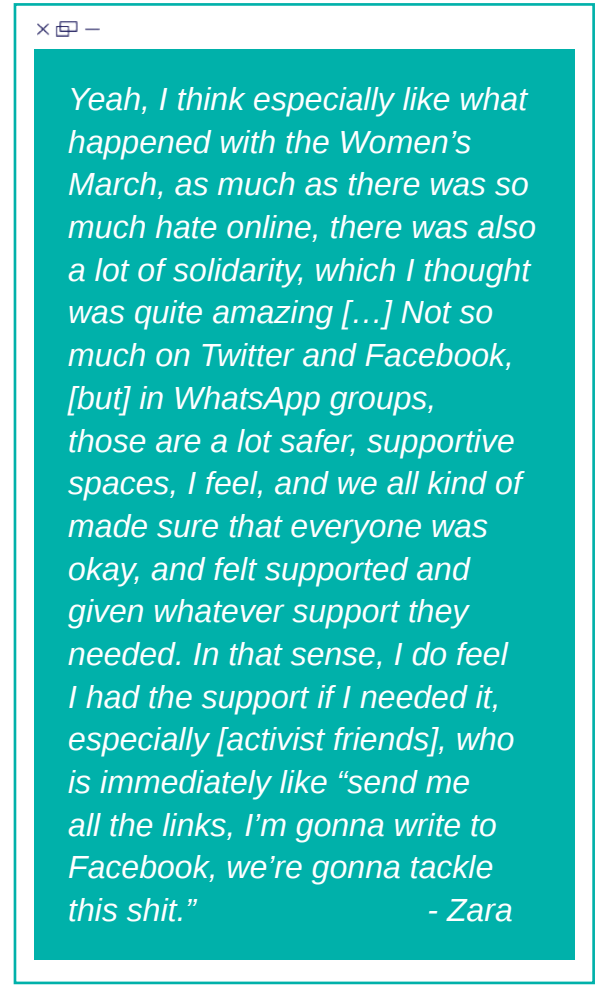
The performance of care and solidarity through private channels needs to be contextualised against the hostile environment in which women and their networks are unable to defend their expressions and discursive space. Suzie tried to push back and defend a friend who was attacked for her speech at the Women's March in 2019 in Kuala Lumpur:

A screenshot of a text message on a teal background. The text is white and reads: "Seeing what happened to [my friend] and I cannot do anything, it made me feel super helpless. I feel like I could've done something better, I should have helped her but I couldn't and it really, really, really is depressing in a way. So I shut down my Twitter for a while because I don't want to face it anymore because I cannot do anything anymore. I've been saying a lot of things [on Twitter], it kinda wore me off in a sense. – Suzie".

Seeing what happened to [my friend] and I cannot do anything, it made me feel super helpless. I feel like I could've done something better, I should have helped her but I couldn't and it really, really, really is depressing in a way. So I shut down my Twitter for a while because I don't want to face it anymore because I cannot do anything anymore. I've been saying a lot of things [on Twitter], it kinda wore me off in a sense. – Suzie

At the height of online mob attacks against Sadia, she privately reached out to two of her friends who were defending her and, as a result of that, they were also harassed on Twitter. She told them, "Enough, you don't have to fight for me," and just let the harassment subside on its own.

In the face of such hostile environments, women were able to forge safe and supportive spaces privately. Yet another exemplification of women's resilience against online gender-based violence and to support one another even when the odds are stacked against them.

A screenshot of a text message on a teal background. The text is white and reads: "Yeah, I think especially like what happened with the Women's March, as much as there was so much hate online, there was also a lot of solidarity, which I thought was quite amazing [...] Not so much on Twitter and Facebook, [but] in WhatsApp groups, those are a lot safer, supportive spaces, I feel, and we all kind of made sure that everyone was okay, and felt supported and given whatever support they needed. In that sense, I do feel I had the support if I needed it, especially [activist friends], who is immediately like "send me all the links, I'm gonna write to Facebook, we're gonna tackle this shit." - Zara".

Yeah, I think especially like what happened with the Women's March, as much as there was so much hate online, there was also a lot of solidarity, which I thought was quite amazing [...] Not so much on Twitter and Facebook, [but] in WhatsApp groups, those are a lot safer, supportive spaces, I feel, and we all kind of made sure that everyone was okay, and felt supported and given whatever support they needed. In that sense, I do feel I had the support if I needed it, especially [activist friends], who is immediately like "send me all the links, I'm gonna write to Facebook, we're gonna tackle this shit." - Zara

Audience curation

Social media algorithms present a novel form of interference to our freedom of expression in which they unilaterally decide the distribution of expressions and information by selecting audience for particular content in automatic and opaque ways. Yet, research participants have shared in interviews steps they have taken to redirect some level of control over the circulation and distribution of their content. Setting their accounts as private is one common and simplest tactic they employ to have better control over the people who will have access to their expression. Other than limiting the number of followers to around 30, Nina also makes

sure her followers on Twitter are friends who share a similar political ideology as hers, which makes it safer for her to express her uncensored self.

The Close Friends feature on Instagram, which allows users to create a list of followers who are permitted to view private content, is another useful tool for women and gender non-conforming persons to protect themselves. While most of the women from the research agreed that the environment on Facebook is a lot more hostile, both Lily and Zainab still find Facebook groups to be a useful feature to connect with queer communities as it accords the administrator the power to manage and regulate membership, an essential step to ensure a safer space for their expression. Yet Zainab had witnessed disparagement towards one transwoman by her community in the same Facebook group when the said transwoman expressed that transwomen can have sexual orientations other than being heterosexual.

Block, mute and report

Block, mute and report are yet another reflexive tactic women adopt, not just as a response to online gender-based violence, but also to moderate their experience on social media. Blocklist, a feature on Twitter that allows users to block multiple people at once and share their lists of accounts they have blocked, is increasingly useful to filter trolls and aggressors. In this sense, blocklists make the work of responding to online gender-based violence more communal and efficient.⁴⁰

By blocking, women and gender non-conforming persons can regain some control over the distribution of their expression on Twitter and participate in public discourse by selectively tuning out their content from reaching the aggressors' community. Most of the women from the research agreed that the use of the Block function has been vital to their expression and how it improves their experiences on Twitter.

The decision by women and gender non-conforming persons to block and mute should also be contextualised against the inadequate/inefficient/ineffective redress mechanism accorded by social media reporting mechanisms and content moderation approach. While most of the women and gender non-conforming persons from the research reported incidents of online gender-based violence to social media platforms, all agreed that their complaints to social media platforms were often not taken seriously, or the platforms replied saying that the contents or accounts did not violate their community guidelines. In this instance, blocking is akin to a cosmetic and temporary solution to the hostile environment on social media that enables the proliferation of online gender-based violence. It also sends a "pass-the-buck" message from these social media on how users should deal with the violence. Just block, mute, etc. However, it does not necessarily mean that the violence stops as we have heard from the research participants. It can continue into various digital spaces as well as into the physical world of those targeted by violence.

40 Geiger, R. S. (2016). Bot-based collective blocklists in Twitter: the counterpublic moderation of harassment in a networked public space. *Information, Communication & Society*, 19(6),787-803. <https://doi.org/10.1080/1369118X.2016.1153700>



CONCLUSION

Understanding social media algorithms and their impact on our freedom of opinion and expression, especially through a feminist lens, requires us to not only closely examine the technical features, which more often than not is out of grasp for many, or deliberately mystified, but to also question the logic that propels the technology. As Shoshana Zuboff puts it, our effort to confront the algorithm begins with the recognition that *we hunt the puppet master, not the puppet*.⁴¹

This paper illustrates that the proliferation of gender-based violence online is an inevitable expression fuelled by social media algorithms when the imbued logic is to drive user engagement for the data economy. The data economy centres itself around the capitalisation of all aspects of human lives and relationships, even the most hideous side of humanity, claiming them as free resources that are simply out there for extraction. These data are then fed into “machine intelligence” and fabricated into prediction products that claim to know who we are and what we will do now, soon and later.⁴²

Closely intertwined with the data economy is the treatment of human’s attention as a scarce resource. The more time and attention we spend on a product, service, post, tweet, video, reels, the more data that can be generated. The ever-flowing information in digital spaces means that companies are constantly courting users’ attention in more competitive ways than ever, to keep their users scrolling, browsing, and engaging with the contents. Such logic privileges incendiary contents that are, among others, racist, misogynist and trans- and homophobic, while ignoring and suppressing feminists’ labour in pushing alternative narratives and

⁴¹ Zuboff, S. (2019). Op. cit.

⁴² Ibid

challenging the dominant and patriarchal discourse. Though women in this paper have demonstrated their resilience in navigating social media's algorithms, the social media companies are still very much in a privileged position to rewrite their algorithms to benefit themselves.

Academic and human rights advocates have actively pushed for alternative digital infrastructures that are rights-based by design, i.e. alternative ways of prioritising content to decrease emotional stimuli so as to offer a slower and calmer environment for networking; to include interventions to question, delay, or limit the reach of hateful comments.⁴³ The feasibility of such design in eliminating online gender-based violence remains to be seen and whether big platforms like Facebook, Twitter and Instagram would ever be redesigned given their core purpose is profit within a capitalist, neoliberal, patriarchal framework.

⁴³ Munn, L. (2020). Op. cit.

