The affective pressures of WhatsApp: from safe spaces to conspiratorial publics

Authors:

Amelia Johns (University of Technology Sydney, Australia, amelia.johns@uts.edu.au)

Niki Cheong (Coventry University, U.K., niki.cheong@coventry.ac.uk)

Word count: 6175

Abstract:

In this paper we bring together media logics (Van Dijck and Poell, 2013; Agur 2019), affordances (McVeigh-Schulz and Baym 2015; Hanckel et al 2019) and affect theory (Lupton 2017) to ask how conspiracy theory moves through WhatsApp groups and moves people through their encounters with these contents toward 'conspiracy thinking' (Onderco and Stoeckel 2020) . Firstly, we draw upon media logic theory to examine the extent to which WhatsApp's architecture, design and technical functions 'steer' users' toward particular communication and behaviours (Van Dijck and Poell 2013). Secondly, we use affordance theory to examine how communities incorporate platform affordances into their tactics to resist, subvert or circumvent institutional power. Finally, we use theories of affect to understand whether the shared emotions and intensities that arise through encounters between digital environments and bodies of users drive experiences that bypass human cognition and representation to create "affective atmospheres" (Lupton 2017, 2). We apply this integrated framework to ask whether WhatsApp's closed infrastructure, end-to-end encryption, and social features such as 'Groups' and the 'forward' button - with the embodied practices of users - shape affective environments which normalise conspiracy thinking.

Keywords:

WhatsApp, Telegram, safe spaces, conspiracy thinking, affect, media logics, affordances

Introduction

In this paper we bring together media logics (Van Dijck and Poell 2013; Agur 2019), affordances (McVeigh-Schulz and Baym 2015; Hanckel et al. 2019) and affect theory (Lupton 2017) to ask how conspiracy theory moves through WhatsApp groups and moves people through these encounters toward 'conspiracy thinking' (Onderco and Stoeckel 2020) . Firstly, we draw upon media logic theory to examine the extent to which WhatsApp's architecture, design and technical functions 'steer' users' toward particular communication and behaviours (Van Dijck and Poell 2013). Secondly, we use affordance theory to examine how communities incorporate platform affordances into their tactics to resist, subvert or circumvent institutional power. Finally, we use theories of affect to understand whether the shared emotions and intensities that arise through encounters between digital environments and bodies drive experiences that bypass human cognition and representation to create 'affective atmospheres' (Lupton 2017, 2). We apply this integrated framework to ask whether WhatsApp's closed infrastructure, end-to-end encryption, and social features such as 'Groups' and the 'forward' button - with the embodied practices of users - shape affective environments which normalise conspiracy thinking.

We draw upon findings from a study conducted between 2016 and 2018 in Malaysia, which included interviews with young political and LGBT activists[1], about their digital, civic and political practices in the lead up to the 2018 Malaysian General Election (GE14). The findings highlight the significance of WhatsApp's end-to-end encryption and closed architecture to the civic and political repertoires of activists and everyday users. This was particularly so within a broader media ecology where censorship, surveillance and manipulation of social media publics fuelled perceptions that WhatsApp's end-to-end encrypted group chats (and opt-in encrypted messaging platforms like Telegram) provided a 'safe space'. However, participants also described other affective responses to the groups they participated in, with the same

features (encryption, closed groups) and social functions (message forwarding) being associated with an uptick of misinformation and conspiracy thinking.

Addressing the theme of the special issue, the findings highlight ambivalent experiences of users of 'dark social' technologies[2]. On the one hand the architecture and affordances of WhatsApp and other messaging platforms allow a range of actors to subvert, resist and evade the logics of surveillance and publicness embedded in social networking platforms and the regimes of state censorship and control that benefit from them. These findings support the work of dark web scholars such as Gehl (2018) and Maddox et al (2016), who critique the 'moral' registers used to understand dark social spaces and the actors who inhabit them. One of the main findings of this study was that 'dark social' technologies enabled activists safety from 'digital light' defined in relation to processes of datafication, state, and platform surveillance. Nonetheless we want to complicate framings that would suggest digital darkness is always productive and progressive; rather the dark social shapes 'affective atmospheres' that are contingent, context-dependent and dynamic, and which can move users from one affective state to another.

Background

In 2018, Malaysia's internet space was declared 'Partly Free' (Freedom House 2018). Freedom House noted– following Malaysia's 14th General Election (GE14) and the first government change in over 60 years– that 'there are hopes that restrictions will lessen'. The restrictions referred to included blocking of online news portals and blogs including *Sarawak Report*, and *Medium*, after they reported allegations of political corruption involving former Prime Minister Najib Razak, who misappropriated funds from the government-operated 1Malaysia Development Berhad (1MDB) strategic development corporation.

The blocking was imposed by the Malaysian Communication and Multimedia Commission (MCMC) that regulates media and communications industries. Over the years, the MCMC has been involved in efforts to censor and manage online information by instructing websites and users to remove 'offensive' content (Johns and Cheong 2019). This is enabled by laws, including the Communications and Multimedia Act 1998 and the colonial legacies embedded in the Sedition Act (in 2015, updated to cover online communication), the now-repealed Internal Security Act, and the Printing Presses and Publications Act – all of which have been used to control speech in the country (Case, 1993).

Nonetheless, these laws did not suppress growing anti-government expressions in an online space dubbed 'the opposition playground' (Cheong 2020). In response, the Najib administration first turned to the mobilisation of 'cybertroopers', that is, covert online actors engaging in political trolling against government's critics. It has been argued that the use of cybertroopers, alongside repressive laws, were part of the administration's efforts to 'exploit and manipulate' network affect to dampen the progress of social movements in Malaysia (Johns and Cheong, 2019, 10).

Another tactic adopted by the government was to steamroll an Anti-Fake News Bill through Parliament pre- GE14, which targeted opposition politicians, media outlets and WhatsApp group admins. The law identified 'fake news' as 'any news, information, data and reports that are whole or partly false' (Fernandez 2019, 179). The vagueness of the definition was likely because, as Lim (2020, 8) argues, 'there has never been a clear legal or public sense of what constitutes 'fake news in Malaysia', considering that 'false and misleading content is already criminalized under other legislation'.

Scholars have offered various terms to capture the types of information and content that can be considered 'fake news' or 'problematic information'. As explained by Jack (2017, 1): 'Some

information … is inaccurate, misleading, inappropriately attributed, or altogether fabricated'. Further, misinformation and disinformation, including rumours and conspiracy theories, describe contents intended to deceive or mislead (disinformation) or are shared without knowledge of the falsehood they contain (misinformation). This interpretation is consistent with the ways participants in this study referred to 'fake news'. Tapsell (2019) has further highlighted 'gossip, rumours and conspiracy' circulating through WhatsApp groups ahead of GE14 as a 'weapon of the weak', which generates feelings of empowerment for citizens otherwise disempowered through normative forms of political participation. We acknowledge this, but also seek to understand how the digital environment itself shapes user experiences, emotions and behaviour. We specifically want to see if 'conspiratorial thinking' – a pattern of thinking (and feeling) where actors 'dismiss authoritative accounts of political events and instead believe in "hidden, malevolent groups secretly perpetuating political plots and social calamities to further their own nefarious goals"' (Oliver and Wood 2014, cited in Onderco and Stoeckel 2020, 1) may be fostered by WhatsApp and Telegram's affective environments.

Theoretical framework

To understand and theorise the ambivalent but often shared thoughts and feeling aroused through participation in WhatsApp groups, in this paper we bring together media logics, affordance theory and theories of affective atmospheres, and examine how they fit within the broader literature on messaging platforms and digital activism.

WhatsApp, Telegram, Signal and other instant messaging applications which have closed architectures; default or opt-in end-to-end encryption for group communication; and affordances which allow media content to be shared between users and groups, have, since their introduction, been identified by activists, journalists, and whistleblowers as platforms which arouse different communication styles and logics to those on social media, as well as a

different set of emotions and feelings. They have been identified as spaces which promote safety, intimacy and authenticity (away from the reputational dynamics and datafication of user-interactions on social networking media platforms, see Van Dijck and Poell 2013; Baym and boyd 2010), and retreat from heavily censored and manipulated mainstream media and social media ecologies. They are also associated with emerging communicative logics, enabling flexible, real-time protest communication and micro-coordination (Ling and Lai 2016, Treré 2015; Lee and Ting 2015).

Treré (2015) argues that too much attention has been paid in the digital activism and social movements literature to the affordances of social media (i.e. Facebook and Twitter) which connect and mobilise 'networked publics', and where publicly available data make these movements easier to analyse at scale. He claims that this focus comes at the expense of research on messaging platforms like WhatsApp, where the closed architecture of the platform prompts different logics focused on internal communication, social cohesion, solidarity and trust building. While Treré uses a media ecologies approach to analyse the movement of activists between the social media 'front stage' (Twitter and Facebook) and 'backstage' (WhatsApp and Messenger), he stresses how use of closed messaging groups doesn't just foster new communication styles but also produces shared feelings of safety, retreat and protection, or 'digital comfort zones' (911).

Lee and Ting (2015) adopt a 'media logics' lens to centre platforms, their technical features and design which steer social practices and communication of activists, while also shaping broader media industries, cultures and institutional practices. Agur (2019) also documents the importance of closed messaging platforms to journalistic information gathering in South-East Asian contexts. Drawing on Altheide and Snow's classic approach, Agur (2019, 181) identifies how 'media formats… organise and technologically structure content and audience

expectations'. He argues that connecting with sources in these contexts require that journalists use messaging platforms, such as WhatsApp, Telegram, Line, and others. This is because of their broad use and popularity, but he also notes how the platform structures logics of 'connectedness' and 'insularity': 'interactions that take place within closed groups and, over time, become isolated in terms of participants and content. Insularity is the result not only of social practices but also of codes, data, algorithms, and interface design' (183). This introduces new ethical considerations for journalists connecting with sources given that users tend to only trust known (internal) contacts and harbour suspicion toward outsiders.

LGBT activists and community members in Southeast Asian contexts also use closed messaging groups in line with these logics. Hanckel et al. (2019) emphasise how closed groups generate feelings of safety, intimacy and trust for young LGBT people. This is because there is considerable risk for them to safely navigate social networked platforms that, by their design, reward visibility and publicness (see also Cho 2018). In choosing closed messaging groups then, they consciously eschew social media logics in favour of safety and 'control of their experience of using the app' (Hanckel et al. 2019). The study draws upon theories of affordances (McVeigh-Schultz and Baym, 2015) to understand queer young people's navigation of digital publicness, and their social media curation strategies which include strategic retreats to closed messaging apps to circumvent risk and create conditions of safety.

Attention has also turned recently to the role that platforms like WhatsApp play in circulating problematic information, primarily misinformation, rumour, conspiracy theory and hate speech. Scholars have focused on the way use of the app increases potential for election manipulation, as in the case of the 2018 Brazilian election (Pereira and Bojczuk 2018) and GE14 in Malaysia (Tapsell 2018). In these cases, mass messaging through WhatsApp groups, enabled by the forward function, and where messages are difficult to trace owing to end-to-end

encryption has allowed inauthentic behaviour to gain a presence among WhatsApp groups discussing the election (Cheong 2020, 78; Pereira and Bojczuk 2018).

Tapsell provides an account of how rumour and conspiracy focused on PM Najib's wife, Rosmah, spread like wildfire through WhatsApp groups in the lead up to GE14, a phenomenon he associated with political parties use of WhatsApp for campaigning, as well as the desire for the 'weak' to seek tools which amplify speech that harms the powerful, while not having repercussions for user communities (Tapsell 2018, 19). In a different register, and not exclusively focusing on messaging platforms, Whyte (2020) provides insight to the role played by reddit, Discord and 4Chan in the election of Donald Trump in 2016. He argues that despite much literature being dedicated to the 'democracy hacking' and manipulation of Facebook and Twitter via bots and false accounts during the 2016 US election campaign, that 'closed, conspiratorial communications' were fundamental to Trump's disinformation machine. Whyte argues that content circulating in these groups and forums is 'characterized by conspiracy-oriented discursive practices that discourage critical thought, "meme-ify" controversial content and encourage hostile rebuttal of external criticism', and that these contents and practices often act as the anchor point and feeder source for algorithmic modes of manipulation.

In this paper we will bring these different scholarly insights - considering closed and default or opt-in encrypted messaging platforms as safe spaces for activists, as well as conspiracy thinking amplifiers – by drawing together theories of media logics, affordances and 'affective atmospheres' (Lupton 2017). We will use theories of 'affective atmospheres' to identify gaps in the aforementioned literature. Rather than locating emotional states and feelings in the rational or conscious interaction of users with digital technologies and affordances, we will reference theories that call attention to how human and non-human interactions in digital environments are 'often felt or sensed by humans entering a place rather than directly observed

or represented', which 'can have profound effects on the ways in which people think and feel about and sense the spaces they inhabit and through which they move ' (Lupton 2017, 1). This framework will be operationalised to analyse informant responses to WhatsApp, and to a lesser extent, Telegram use in 2016 and the lead up to GE14 (2018).

Methods

The data informing the study is drawn from two field trips to Kuala Lumpur between 2016 and 2018, a time of dramatic upheaval in Malaysia's political structure (Johns and Cheong 2019). This was also a time when participation in encrypted chats on Messenger, Telegram, and WhatsApp was recognised as a game changer, subverting Malaysian authorities' efforts to curb political activism through surveillance and laws.

The sample for this paper consisted of 29 informants of Malaysian-Chinese ethnicity, aged 18-24, who were recruited using targeted as well as convenience sampling. Advertisements were placed in known activist social media (including by invitation to closed WhatsApp groups), as well as university student association Facebook groups and social media. Malaysian-Chinese young people were purposively recruited into the study owing to media accounts that this community were crucial to online movements and activism to change Malaysia's leadership (Tapsell 2018).

In keeping with best practice working with digital communities who resist intrusion (Swart, Peters and Broersma 2018) WhatsApp and Facebook group admins were contacted and permission sought to monitor closed communications. In addition, semi-structured interviews of 60-minutes duration were undertaken with 21 informants in 2016 using techniques to open up interviews beyond structured questioning, i.e. use of video and screen capture to elicit responses from participants as they scrolled through their feeds, demonstrating daily social

media routines and habits. A further 8 informants were interviewed in 2018, including follow up interviews with 4 original participants.

Findings

The safety of WhatsApp

Among activists we interviewed half were LGBT activists whose focus was on more than changing the political structure, though this was also a concern. Because of social stigma attached to this community, combined with fears that communication on public social media might incur the surveillance and legal machinery of the state, the 'default publicness' (Cho 2019) of platforms like Facebook was eschewed, and WhatsApp or opt-in encrypted chats on Telegram were preferred. One informant also discussed 'secret groups' on Facebook, and on Messenger:

> Usually secret groups are highly sensitive or things that you don't want other people who are not in the group to know… closed groups are sort of like I don't mind being in the group but I don't want people to read unless they are in the group. Open is public. (Phillip, LGBT activist, 21).

Julian, a student activist, expressed a preference for Telegram but identified WhatsApp as a more popular platform in Malaysia owing to the way messaging apps were marketed and taken up in this context (Ling and Lai 2016); he particularly discussed the politics surrounding encryption, with Facebook's entry into this market in 2016:

> Informant: everyone's on 3 different chat apps at once, because Singapore and Malaysia got hit 3 ways at the same time. So they've had WhatsApp for a while and then WeChat came in from China and then Viber…now there's like Telegram for paranoid people.

> Interviewer: So why is Telegram for paranoid people? WhatsApp's got end-to-end encryption too?
>
> Informant: It does now. And at the time they were all saying, 'Oh Facebook's going to buy WhatsApp, we're all going to be screwed, they're going to suck in our phone numbers and our profiles against our will and we'll have to delete everything manually.' So I was freaking out...except our worst fears have not come to pass somehow and they've also introduced encryption. So we are at least safe from the government, but not necessary from Facebook (Julian, 24).

While Julian describes concerns regarding surveillance by digital platforms, nonetheless the end-to-end encryption guarantee was identified as a form of protection from government surveillance. This affordance was discussed in the context of a spate of arrests that had caused fear among politically engaged young people in 2016, a period of time when the Sedition Act and other legal measures were being used to curb contentious speech:

> People are being arrested just because they posted something that insults the government or Prime Minister. A few weeks back, the founder of Seksualiti Merdeka and a few of his friends were arrested. (Steven, 23).

As scholars have noted, the growth in popularity of encrypted by default or opt-in encrypted messaging applications, particularly WhatsApp and Telegram, in Malaysia, has, in part, been due to a tightening of the regulatory environment for social media participation and speech since the 2013 election. This is also reflected in non-activist informant accounts, where topics or postings that could be considered at risk of coming under scrutiny by MCMC led them to seek the most secure and protected spaces they could.

Ben (21) was on numerous political WhatsApp groups, but never engaged in political chat on Facebook for fear of legal repercussions and reputational damage for him and his family. He cited the Sedition Act as a reason why many Malaysian-Chinese people felt unsafe on open social networks, as he felt it was used by the government to suppress criticism of special *Bumiputera* rights in Malaysia and increasing Islamisation of the public sphere, both topics considered contentious and incurring claims of sedition:

> It's an invisible box, and we call that box the Sedition Act … it is omnipresent, always in the background waiting for you to make some sort of statement that you can get hauled away for. And I think a lot of people do live in perpetual fear of that (Ben, 18).

Rhys, a law student, who also discussed being on multiple WhatsApp groups for political discussion, described WhatsApp as 'safer' than social networking platforms like Twitter, where even if account settings were set to private, the networked architecture of Twitter made it easier to access information:

> I think [WhatsApp] was somewhat safer... safe for me to discuss politics, rather than discussing it publicly … because there are always people watching even if you've set your settings to private. (Rhys, 24).

The latter comment referred to a friend of his who was doxed on Twitter despite having a private account. Some activists felt a compulsion to share their political views in Facebook status updates and on Twitter or Instagram feeds, so as not to be silenced; but they also discussed the importance of encrypted group chats in a manner which reflected Treré's analysis of WhatsApp as a social media 'backstage' (Treré 2015). Billy, who worked for an NGO seeking an end to political corruption and Simone, who volunteered for a WhatsApp group who

provided legal assistance to citizens arrested under sedition charges, spoke about the types of communication that these platforms afforded:

> It's one of our primary communication methods internally, for staff and committee members. To mobilize for example, we have a whole bunch of groups. So, the last bi-elections, we form WhatsApp groups for the volunteers; to disseminate information, or mobilize people for certain things, it's just a matter of mass communicating (Billy, 26).

> Like when someone gets arrested for any posting, they'll post it here and then they will contact that person's family or that person directly, and offer legal assistance. (Simone, 19).

In thinking through the experiences informants had of using WhatsApp and Telegram, the architecture and design of these platforms, and particularly the feature of end-to-end encryption seemed to make this a preferred option for privacy conscious activists and everyday users, and mention was often made of it being 'safer' than public social media communications (Agur 2019). Of course, this was also informed by broader media ecologies where surveillance of networked publics and laws to censor speech and arrest users made retreats to the social media "backstage" more urgent.

The type of communication logics steered by closed, encrypted chat also reflect Treré's discussion of WhatsApp as a 'digital comfort zone' (Treré 2015) where the focus is on internal cohesion building, intimacy and trust. Despite Billy arguing that WhatsApp is used by organisers to mobilise volunteers for specific campaigns, shadowing the communicative logics more commonly associated with social media platforms; the dominant take away from informant responses was the feeling of 'safety' that encrypted chats offered, even if it was ephemeral and in some cases illusory as the next section reveals.

Spies and rumours

Despite WhatsApp providing assurances to activists, journalists and citizens that their conversations were able to be kept safe in the 'dark', in 2016-2018, high profile cases where WhatsApp group conversations were leaked or compromised, raised alarm and began to replace perceptions of WhatsApp as a 'safe space'. In one case the arrest of a 76-year-old 'uncle' in a WhatsApp group where he shared an image that caused 'offense' to the Prime Minister, reverberated through activist communities:

> Recently one man was arrested if I'm not mistaken ... for posting things on WhatsApp that were insulting to the Prime Minister (Steven).

The broad assumption seemed to be that government informants, or police, had infiltrated the group by requesting invitation under a pseudonym, making end-to-end encryption redundant. Once in they were able to share screenshots to authorities:

> Because even though no matter how secure it is people can simply screenshot and give it to the police or give it to the administrative of the school (Hai Yang, 22).

> I wouldn't give it [political opinion] in WhatsApp or any recorded setting. Because it's not good to keep records… words become very sensitive.… There are news that police tend to catch people who are spreading the rumours on WhatsApp (Anita, 22).

The findings showed that human-centred methods of collecting and sharing information outside of the closed chat could not be stopped by encryption guarantees, despite their aura of impenetrability. Julian also claimed that several activist groups he belonged to had been compromised by spies:

> We would try to use WhatsApp and Telegram encrypted platforms in order to avoid espionage from government, but of course the technology is only as strong as the

> human link. So there were cases that came up in the year previous to the election where there were student groups that were affected where the government would simply insert one of their agents into the group … You have no protection from the encryption at that point. (Julian).

Nonetheless, Given that activists and pro-opposition supporters held a world-view where they were suspicious of government and public institutions (Tapsell 2018) it was not difficult to see how real though isolated cases could be escalated into a conspiracy that spies lurked everywhere, particularly given the logics of 'insularity' and suspicion toward outsiders noted by Agur (2019). As Julian admitted, much of the rumour circulating in WhatsApp groups leading up to GE14 began with distrust of government then anxieties grew in the dark, and echoed through WhatsApp's closed though connected groups.

This is a key point that warrants attention. For insular and closed, though connected (as Agur suggests) WhatsApp publics, lack of platform moderation (Matamoros-Fernandez 2020) combined with the social function of the forward button enabled rumours to be forwarded easily. Julian recalled how the groups he was plugged into in the lead up to GE14 reflected these types of mediated anxiety and paranoia:

> I was plugged into a group of diaspora members in Melbourne and ... every evening there would be one coming in. There would be like, 'this person has been corrupt – how much money they have embezzled – they are doing this – they are scheming this'. Like it switches an opinion in favour of certain political figures or parties - no-one really knows who is sending out these messages. No-one really knows if they are true or not. The people believe it because of course the government is visibly corrupt (Julian).

This analysis shows that beyond the strategic and rational retreat of actors to WhatsApp and Telegram – where the architecture and affordances of these platforms are understood to shield users from the digital light – in some contexts feelings of suspicion toward outsiders and anxiety regarding 'spies' or fabricated claims about politicians can also grow. This, combined with the forward function and the ease with which contents can be shared between groups, facilitates affective states in some communities of users, where pre-formed suspicions interact with the 'haptic' qualities of the platform, primarily the forward button, driving forms of anxiety that are 'more than human' and which can help gain legitimacy and normalise belief that secret plots are everywhere.

Fakery and conspiracy

The belief in WhatsApp as a 'safe space' was further challenged by informants interviewed after GE14 – including those who in 2016 expressed a view that WhatsApp was a safe and progressive discursive space, but who in 2018 expressed concerns that WhatsApp had become a source of 'fake news':

> I am very cautious about information on WhatsApp. I did have a family group … … cousins and relatives sharing like – oh you could say like fake photos (of politicians) (Anita, 24).

Cassie, a journalist who in 2016 used a WhatsApp group for work but felt uneasy about the private groups she was in— because she felt that the closed architecture had a tendency to produce echo chambers among her social group (predominantly anti-government)— had her feelings confirmed in 2018 when she claimed that the circulation of misinformation and rumour was rife, with much of it seeming to suggest government conspiracies were afoot. In the interview, Cassie showed me a message she had received which insinuated that the government were in 'secret talks' for PM Najib to stand down to try to save votes for BN (see Fig 1). She

particularly described the amplifying function of the forward button, which made it easy for anxious or non-digitally literate users to forward content, with messages getting ‘forwarded and forwarded’ (see Fig 2):

[Figure 1 goes here]

Fig 1. WhatsApp chain mail

Here, the haptic function of the forward button and the emotive contents of some of the messages detected by informants as ‘fake news’ resonates with Lupton’s work on ‘affective atmospheres’ created in the digital health space where Lupton argues that: ‘Affective atmospheres are shaped by their multisensory properties: how spaces and places are physically encountered via their visual, haptic, aural, olfactory and taste properties is central to the feelings they generate’, which in relation to digital health technologies, Lupton likens to wearables which vibrate, buzz or ‘tap’ users, prompting affective reactions and states of feeling such as happiness or anxiety (Lupton 2017, 1, 7).

The content of the messages discussed here also generated affective responses. The messages themselves, according to informants, appeared in ‘long text message’ format often including capitalised characters, and had instructions that warned participants that an emergency was unfolding and for their own or their loved one’s safety they needed to forward it on. This was referred to as being like ‘chain mail’.

[Figure 2 goes here]

Fig 2. Chain mail message

Simon (22), a non-politically active informant spoke of several examples of messages he received through WhatsApp family groups:

[Figure 3 goes here]

Fig 3. Alarmist message sent through Family WhatsApp group

He acknowledged the way that these contents manipulated people like his father, who shared not just because of panic, but because of love and care for his family and a feeling he was helping. But even in this scenario Simon claimed that the suspension of rational thought and the 'affective atmosphere' of WhatsApp also competed with quite rational behaviours, i.e. recognition that the messages were hidden from authorities and largely unmoderated. As a result he felt his dad didn't think twice about sharing:

> They don't kind of think they're being watched at all or it could possibly come back to them … (Rhys)

Conclusion

As discussed by informants, the closed architecture and guarantee of end-to-end encryption has made WhatsApp an essential 'backstage' for activists, enabling communities to resist and subvert state based practices of surveillance and control of public conversations on social media. Informants also supported claims that communicative logics of 'insularity', trust and authenticity associated with the closed architecture and the encryption guarantee of WhatsApp and Telegram (Agur 2019; Johns 2020) differed from social media logics of connectivity, publicness, visibility, datafication (van Dijck and Poell 2013) and 'reputational management' (Baym and boyd 2012). These newer logics were commonly identified with increased feelings of safety and trust among users and between users and the platform. Nonetheless, some informants discussed more ambivalent feelings even as early as 2016. By 2018, in the lead up to GE14, this ambivalence had increased with some informants believing that pressures to determine validity and quality of information on more public-facing social media fell away on

WhatsApp, with lack of moderation and inability for authorities to access conversations leading problematic information to circulate.

The logic of insularity (Agur 2019) is interesting to the current study as it explains the retreat of activist and other political communities to these platforms, where concerns about surveillance and being subject to 'moderating mechanisms' by corrupt anti-democratic governments (for example via sedition laws and communications and multimedia act) structures retreats to 'safe spaces' where members can avoid censorship, allowing a free exchange of ideas critical to democracy (Treré 2015). But the insularity of members within a closed 'socio-technical system', also *structures* practices that can lead users to only trust insiders and believe mainstream news and 'official' accounts to be inauthentic and fake. This allows problematic information that contradicts official information, including rumour and conspiracy, to flow more freely through WhatsApp groups.

We also gave more consideration to how users connect, create and share content differently in closed and encrypted messaging platforms but also to consider how this contributed toward shared feelings and perceptions. We drew on theories of affect and 'affective atmospheres' (Lupton 2017) to consider how, away from the institutional light and visibility of social media platforms, information that is false, often deliberately so, can be shared more easily. For example, receiving information from a known, trusted contact increases our willingness to believe its validity and to forward on. If the information creates a sense of anxiety and urgency this can also move users to share and suspend rational judgement. Informants in this study suggested that the emotive content of messages they identified as 'fake news', i.e. warning of dark, secretive plots that would harm society, combined with the haptic qualities of the forward function and the assurance from end-to-end encryption generated feelings that moved users to anxiously share. The constant attention being paid to these contents which were 'forwarded

and forwarded' between groups were shown in some informants' view to reinforce and normalise belief in the conspiratorial content of the messages.

While more research is required, we believe that affect theory can help us to account for these contingent, and sometimes quite unexpected shifts in the shared, embodied feelings and beliefs that arise from users' encounters with dark social technologies, encouraging researchers to avoid framing user behaviour and practice as being influenced either purely by the technologies or the rational decisions of actors. Rather in this paper we have considered how affective atmospheres account for the 'more than human' interactions that occur beyond these limiting dyads.

Acknowledgements

We would like to acknowledge Deakin University, who provided seed funding for the project, our informants and the organisations who assisted with recruitment.

Declaration of Interest Statement

No potential conflict of interests were reported by the authors.

References

Agur, C. 2019. "Insularized Connectedness: Mobile Chat Applications and News Production." *Media and Communication*, 7 (1): 179–88. doi:10.17645/mac.v7i1.1802.

Baym, N. and d. boyd. 2010. "Socially Mediated Publicness: An Introduction" *Journal of Broadcasting & Electronic Media*, 56 (3): 320–329.

Case, W. 1993. "Semi-Democracy in Malaysia: Withstanding the Pressures for Regime Change." *Pacific Affairs* 66 (2): 183–205. doi.org/10.2307/2759366.

Cheong, N. 2020. "Disinformation as a Response to the 'Opposition Playground' in Malaysia."

In *From Grassroots Activism to Disinformation: Social Media in Southeast Asia*, edited by Aim Sinpeng and Ross Tapsell, 63–85. Singapore: ISEAS Publishing.

Cho, A. (2019). "Default publicness: queer youth of color, social media, and being outed by the machine". *New Media & Society* 20 (9): 3183–3200. doi.org/10.1177/1461444817744784.

Fernandez, J. 2019. "Malaysia's Anti-Fake News Act." *Pacific Journalism Review* 25 (1 & 2): 173–92.

Freedom House. 2018. "Freedom of the Net 2018: Malaysia." Accessed 18 April 2021. https://freedomhouse.org/country/malaysia/freedom-world/2018.

Gehl, R. W. 2018. *Weaving the dark web: legitimacy on Freenet, Tor, and I2P*. Cambridge, MA: The MIT Press.

Hanckel, B., S. Vivienne, P. Byron, B. Robards, and C. Churchill. 2019 "'That's not necessarily for them': LGBTIQ+ young people, social media platform affordances and identity curation." *Media, Culture & Society*. 41 (8): 1-18. doi.org/10.1177/0163443719846612

Hopkins, J. 2014. "Cybertroopers and Tea Parties: Government Use of the Internet in Malaysia." *Asian Journal of Communication* 24 (1): 5–24. doi.org/10.1080/01292986.2013.851721.

Jack, C. 2017. *Lexicon of Lies : Terms for Problematic Information*. https://datasociety.net/output/lexicon-of-lies/.

Johns, A, and N. Cheong. 2019. "Feeling the Chill: Bersih 2.0, State Censorship, and 'Networked Affect' on Malaysian Social Media 2012–2018." *Social Media + Society* 5 (2): 1–12. doi.org/10.1177/2056305118821801.

Lee, A.Y. L., and K.W. Ting. 2015. “Media and Information Praxis of Young Activists in the Umbrella Movement.” *Chinese Journal of Communication*, 8 (4): 376–92. doi:10.1080/17544750.2015.1086399.

Lim, G. 2020. *Securitize/Counter-Securitize: The Life and Death of Malaysia’s Anti-Fake News Act*. https://datasociety.net/library/securitize-counter-securitize/

Ling, R, and C. H. Lai. 2016. “Microcoordination 2.0: Social Coordination in the Age of Smartphones and Messaging Apps.” *Journal of Communication* 66 (5): 834–56. doi.org/10.1111/jcom.12251.

Lupton, D. 2017. “How Does Health Feel? Towards Research on the Affective Atmospheres of Digital Health.” *Digital Health, no.* 3 : 1-11. doi.org/10.1177/2055207617701276.

Maddox, A, M. J. Barratt, M Allen, and S. Lenton. 2016. “Constructive Activism in the Dark Web: Cryptomarkets and Illicit Drugs in the Digital ‘Demimonde.’” *Information, Communication & Society* 19 (1): 111–26. doi.org/10.1080/1369118X.2015.1093531.

Matamoros-Fernández, A. 2020. “‘El Negro de WhatsApp’ meme, digital blackface, and racism on social media. ” *First Monday* 25 (12). doi.org/10.5210/fm.v25i12.10420.

McVeigh-Schultz, J, and N. K. Baym. 2015. “Thinking of You: Vernacular Affordance in the Context of the Microsocial Relationship App, Couple.” *Social Media + Society* 1 (2): 1-13. doi.org/10.1177/2056305115604649.

Onderco, M, and F. Stoeckel. 2020. “Conspiratorial Thinking and Foreign Policy Views: Evidence from Central Europe.” *Journal of Elections, Public Opinion and Parties*, 1–15. doi.org/10.1080/17457289.2020.1814309.

Pereira, G., and I. Bojzcuk. 2018. *Zap Zap, Who’s There? WhatsApp and the Spread of Fake News During the 2018 Elections in Brazil – Global Media Technologies and Cultures Lab*. Accessed 19 Jan. 2021. http://globalmedia.mit.edu/2018/11/09/zap-zap-whos-there-whatsapp-and-the-spread-of-fake-news-during-the-2018-elections-in-brazil/.

Swart, J., C. Peters, and M. Broersma. 2018. “Shedding Light on the Dark Social: The Connective Role of News and Journalism in Social Media Communities.” *New Media & Society* 20 (11): 4329–45. doi.org/10.1177/1461444818772063.

Tapsell, Ross. 2018. “The Smartphone as the ‘Weapon of the Weak’: Assessing the Role of Communication Technologies in Malaysia’s Regime Change.” *Journal of Current Southeast Asian Affairs* 37 (3): 9–29. https://journals.sub.uni-hamburg.de/giga/jsaa/article/view/1146/1153.

Treré, E. 2015. “Reclaiming, Proclaiming, and Maintaining Collective Identity in the #YoSoy132 Movement in Mexico: An Examination of Digital Frontstage and Backstage Activism through Social Media and Instant Messaging Platforms.” *Information, Communication & Society* 18(8): 901–15. doi:10.1080/1369118X.2015.1043744.

Van Dijck, J., and T. Poell. 2013. “Understanding Social Media Logic.” *Media and Communication* 1(1): 2–14. doi:10.12924/mac2013.01010002.

Whyte, C. 2020. “View of Of Commissars, Cults and Conspiratorial Communities: The Role of Countercultural Spaces in “Democracy Hacking” Campaigns.” *First Monday* 25(4). doi.org/10.5210/fm.v25i4.10241.

[1] Lesbian, Gay, Bisexual, Transgender. This was the most commonly used terminology among participants.

[2] 'dark social' refers to social traffic or shares that do not contain information referring back to a source, and can’t be traced. So, unlike shares from social media platforms, dark social refers to sharing of contents through private channels, like email, SMS and messaging platforms.