Surveillance does not Equal Safety: Police, Data and Consent on Dating Apps

Zahra Stardust[1,2], Rosalie Gillett[1,2] and Kath Albury[1,3]

[1] Australian Research Council Centre of Excellence for Automated Decision-Making and Society

[2] Digital Media Research Centre, Queensland University of Technology

[3] Department of Media and Communication, Swinburne University of Technology

Key words: Dating apps, user safety, sexual consent, online harm, gendered violence, privacy, surveillance

Abstract: As dating apps continue to receive pressure from civil society, media and governments to address a range of safety concerns, technology companies have developed and deployed a spate of new safety features. Taken together, these features rely upon increased surveillance and partnerships with both technology start-up companies and law enforcement agencies proposed as responses to sexual harassment and abuse. In this article, we draw on empirical accounts of app use—and popular media reporting—to problematise commonsense assumptions about dating apps, safety, technology, policing and surveillance. Where so-called safety features involve increased surveillance and techno-carceral solutionism, there is potential to make users *less* safe—particularly for app users who are marginalised or stigmatised on the basis of their race, sexuality, gender, health status, employment or disability. Instead of the impetus to 'datafy' consent by documenting evidence of sexual transactions, or to monitor users by sharing data with police, we argue that a more effective approach to safety must extend the notion of 'consent culture' to encompass a consent-based approach to collecting, storing, and sharing user data – including seeking consent from users about how and whether their data is sold, monetized or shared with third parties or law enforcement.

Introduction

Dating apps have recently been under increased scrutiny for their role in facilitating harassment and abuse. In response, dating apps have introduced various automated systems in attempts to mitigate abuse and address user safety needs, with user verification, increased surveillance, and partnerships with both external 'consent apps' and law enforcement agencies proposed as responses to sexual harassment and abuse. While some dating apps are making sincere attempts to prevent and address abuse, they expand opportunities for social monitoring on the part of governments, police and corporations. Where so-called safety features involve increased surveillance, there is potential to make users *less* safe – and this is particularly the case for app users who are marginalised or stigmatised on the basis of their race, sexuality, gender, health status, employment or disability. Dating apps elicit data that is susceptible to both vertical and lateral surveillance: they gather and share intelligence with law enforcement at the same time as providing spaces where users can police stigmatised identities. Like other social media apps and platforms, many popular dating apps are constrained by business models that prioritise 'connection' and data extraction, infrastructures that do not fully account for sexual or gender diversity, and logics that do not appreciate the impacts of structural oppressions. Many appear to lack a comprehensive framework for conceptualising consent, focusing predominantly on the risks of sexual assault and inadequately on more insidious and systemic forms of harm. In implementing such 'safety' initiatives, dating apps operate to unearth substantial quantities of monetizable data, while also consolidating a heteronormative socio-sexual order and strengthening a carceral state.

In this article, we draw on public announcements by app developers, as well as news reporting on app features, to reflect on the ways 'safety' has been framed as 'risk' in public conversations about apps. Media and tech conversations about safety continue to focus on new forms of surveillance to mitigate risk, such as offering police access to dating app data, encouraging users to document their sexual consent, verifying users via photo identification, and introducing automated speech detection tools. These public statements and initiatives are contrasted with the findings of two recent empirical studies of Australian dating app users' experiences and perceptions of dating apps – one which asked women about their experiences of intimate intrusions on Tinder, the other asking app users of diverse sexualities and genders to explain the features of app design and app culture that made them feel safer or less safe. These studies indicate that users have vast and subjective perceptions of 'safety' on dating apps, including different views on what constitutes 'red flags', alternative means of vouching for users, varying attitudes towards negotiating consent, and specific needs in relation to data privacy. Through this approach, we seek to problematise the assumption that increased surveillance and the presence of law enforcement on apps (which we associate with broader tendencies towards technical and/or carceral solutionism) will equate to safer user experiences. We argue that these tendencies – such as the impetus to 'datafy' consent by seeking to document evidence of sexual transactions, to verify users through photo recognition, or to share proprietary data to inform policing – will not necessarily protect dating app users. Instead, we conclude by suggesting an alternative approach to safety, that extends the notion of 'consent culture' to encompass a consent-based approach to collecting, storing, and sharing user data.

Dating apps provide an apt object for analysis because they collect intimate, sensitive and stigmatized user data. They therefore offer insights into the broader phenomenon of social media and carceral surveillance in addition to the movement for ethical data governance of

digital technologies. Existing research demonstrates how "[s]ocial media surveillance reduces individuals' control over the information they disclose about their attributes in different social contexts, often to powerful actors such as the state or multinational corporations" (Brown, 2014, 2) and how "emergent technologies augment the police's control of their public visibility and that of the social world" (Walsh and Connor, 2018, 1). Feminist scholars have illustrated how surveillance practices serve to "normalize and maintain whiteness, able-bodiness, capitalism, and heterosexuality" (Dubrofsky and Magnet, 2015, 6) and are linked to the "burgeoning prison industrial complex" (5). While some literature addresses the legal risks for queer dating app users (Green, 2019) and the prevalence of sexual racism towards Indigenous dating app users (Carlson, 2020), there is a distinct gap in research about the potential implications of police surveillance on dating apps, especially in terms of how it may impact user safety. And yet, social media's affordances continue to augment police surveillance and intelligence-gathering by expanding their reach to automated and predictive forms of monitoring (Rigot, 2022; Brayne, 2021).

As regulators in multiple jurisdictions seek to identify the appropriate limits of the data subject, establishing rights for users to access and delete their own data or to be 'forgotten', there is an increasing movement for platform accountability and transparency around the collection, use, sale and sharing of user data. However, as Wark (2019) observes, the collection and circulation of data – or information is a central organizing feature of contemporary power relations. While it may not be possible to design 'safe' or ethical dating apps within the context of what is commonly known as surveillance capitalism, we note that users consistently navigate oppressive, exploitative, extractive and otherwise imperfect systems in order to meet their needs. Regardless of whether users can entirely escape 'datafication', we argue that data systems ought to be re-made: that users ought to have more nuanced choices about if, when, how, where and why their data is collected, used and shared,

and clearer information about how their data are being monetized and analyzed (and by whom). The trifecta of police, surveillance and identity verification does not necessarily make dating apps safer, but rather adds further risks for users to navigate.

Safety and risk in dating discourse

As platforms that facilitate both online and in-person encounters, dating apps have been under civil society, media, and state pressure to address a range of safety concerns. In 2017, Tinder rolled out 'Reactions'—animations, which mimicked 'real life' reactions people might have offline, imported into Tinder's messaging service, including throwing a martini in someone's face and an eye roll. Tinder touted the tool as an effective way to call out men's 'douchey' behaviour (Tinder, 2017). The company launched Reactions as part of the platform's broader 'Menprovement Initiative' to "promote a positive community and make it easier than ever for […] users to take action on Tinder" (Tinder 2017), As part of the 'Menprovement Initiative', Tinder also introduced an improved reporting mechanism and updated their community guidelines, suggesting that Reactions accompanied a broader campaign designed to make the platform safer for users.

While Reactions was short-lived, the initiative nevertheless provided an insight into how the platform understood safety at that time. More recently, Tinder has developed and deployed several new safety features and updates including a new Safety Center and automated speech detection tools (Tinder, 2021a, 2021b), indicative of the industry's steps to address abuse and in-app culture. Shortly after Tinder's most recent safety developments, the dating app Bumble announced its partnership with the survivor-led non-profit organisation Chayn to deploy Bloom—a remote trauma support service for survivors of gender-based violence (Naidoo, 2021). Although the effectiveness of dating apps' new safety initiatives remains

unclear, these initiatives suggest that the platforms are doing more to centre users' needs and are taking safety more seriously.

In a recent study of popular media reporting related to Tinder, Bumble and Grindr, Albury et al. (2020) observe a tendency for both journalists and expert commentators to frame app use as intrinsically risky, drawing causal links between app use and poor sexual health and mental health outcomes and crime (ranging from blackmail to sexual assault). The authors note that these links between app use and risk are foregrounded even in cases where there is no clear evidence of correlation between apps and harmful events. Given that non-technologically-mediated intimate relationships can also be associated with these outcomes, it is possible to argue that app use is 'at least as safe' - or risky - as any other form of dating. Additionally, some dating app users have reported feeling 'safer' as a direct result of app use (Albury et al., 2019; Byron, 2020).

While an exhaustive interrogation of the notion of 'dating safety' is beyond the scope of this paper, we begin by problematising the assumption that dating app use is intrinsically risky. Möller and colleagues (2006) observe that while an understanding of safety is intrinsic to the concept of risk, the former has been significantly undertheorised when compared to the latter. As they note, safety is often represented as "the inverse of risk" – particularly in technical usage (Möller, 2006: 419). However, most risk-modelling assumes that risk can only be reduced as opposed to eradicated, meaning that absolute safety is unattainable. In these models, the notion of safety makes no sense without an understanding that some level of risk is present.

Möller et. al. further note that multiple studies correlate subjective judgements of risk with perceived levels of control--offering the example of a person who feels 'safer' when they carry a gun, even though statistical evidence suggests that that gun-owners are more likely

than non-owners to be killed or injured. Consequently, they conclude that while safety cannot be defined or understood in objective terms, it cannot be understood as purely subjective either. Instead, the authors propose a third position – that of *intersubjective safety* – in which varying scenarios or states of being are evaluated in relation to one another. That is, rather than defining one scenario as safe and another as risky, it is more appropriate to consider whether one scenario is 'at least as safe as' another, given the relative probabilities of dangerous outcomes.

In a linguistic analysis of a corpus of 'everyday' (and journalistic) uses of the terms risk, safety and security, Boholm and colleagues (2016: 330) note that there "are tendencies in ordinary language to associate safety with unintentional harm and security with intentional harm." That is, the term safety is most often used in relation to discussions of accidents, and is associated with protective technologies such as safety belts, or safety nets. In contrast, security is more often associated with deliberate attempts to harm (such as data breaches or military action). They note too, that forms of vernacular usage often differ from academic uses of these terms. Given that risk is often desirable in everyday terms--for example, as Sonia Livingstone (2015: 334) notes, it can equate with opportunity--the authors provocatively suggest that where academics and others seek to convey a technical (as opposed to vernacular) understanding of 'risk', they are better off using the term 'expected damage.'

So, what kinds of 'expected damage' are implied in conversations and interventions relating to dating apps and safety? The issue of data security looms large in this context. If we understand app users' safety in intersubjective terms, it is clear that mandatory verification and authentication via legal identity documents does not necessarily keep users safe—rather, it can put some at greater risk (boyd, 2012). When 'cheater's' dating service Ashley Madison

had its database hacked, personal details including usernames, passwords, home addresses, email addresses, credit card information, GPS coordinates and sexual fantasies of around 37 million users were released (Cross, Parker and Sansom, 2019). Additionally, where apps collect sensitive information about their users' sexual health status, gender identity and sexual preferences, poor privacy policies can put users at risk of stigma, entrapment, arrest, and deportation (Albury et al. 2017). The increasing interest of law enforcement agencies in dating app data sits within a broader project of escalating police surveillance, fueled by platform-verification processes. Consequently, the trend towards 'real name policies', requiring users to upload their identities, or authenticate their identities using verified social media log-ins, facial recognition technologies or photo-matching has been deeply contested by members of marginalised communities (Kornstein, 2019; van der Nagel, 2015).

Interdisciplinary approaches to safety and risk on apps

In this paper, we write from interdisciplinary backgrounds that span digital media, criminology, law and gender and sexuality studies, and draw upon our current position as researchers of automated technologies. The authors have elsewhere interrogated how platforms frame their safety interventions in order to justify investment and upscale of automated technologies (Gillett, Stardust and Burgess, forthcoming), and have documented negative impacts of surveillance technologies (Blunt and Stardust, 2021), entrapment, policing (Stardust et al, 2021) and carceral approaches (Stardust and Caldwell, 2022) upon marginalised communities. In this article, our reflections on app user perceptions of safety draw upon two Australian empirical research projects that examined the experiences of dating app users of various genders and sexualities in relation to harm, safety, and consent (Gillett, 2019; Albury et al 2019, 2020). Gillett's (2019) research investigated women's

'everyday' experiences of 'intimate intrusions'—a term coined by Elizabeth Stanko (1985: 1) to describe men's behaviour that women experience as "intimidating, threatening, coercive or violent." Gillett's (2019) study, however, built on this definition to include men's behaviour that makes women feel uneasy, uncomfortable, or unsafe in the context of their Tinder use. The study used a mixed methods approach, drawing on semi-structured interviews, complemented by interview 'scroll backs' (Robards et al. 2012), and a walkthrough (Light, Burgess and Duguay 2016) of Tinder to understand the platform's features and functions, and environment of expected use. Participants in the study included seventeen 18-30-year-old women who were living in Brisbane (a relatively small Australian city) and had experienced what they described as intimate intrusions by men during their Tinder use. To protect the identities of the participants, Gillett (2019) did not collect their demographic data.

Albury's research investigated the experiences of dating and hook-up app users of diverse sexualities and genders, aged 18-35, in which participants were invited to reflect on the elements of app design and app cultures that made them feel safer or less safe. This project was undertaken in partnership with two non-government sexual health organisations, and received both university ethics approval and ethics clearances from the partner's in-house ethics committees. It deployed a mixed-methods approach, including iterative consultations with project reference groups; a review of media reports featuring dating apps (Albury et al. 2020); a survey of Australian app users (n=382); a series of workshops with urban and regional app users drawing on creative and participatory methods (n = 32), and semi-structured interviews with app users in two states (n= 24). Participants in this study reported using a wide range of dating and hookup apps, including Tinder, Bumble, HER and Grindr. Participants explained the ways they weighed up safety and risk in dating app culture in terms of designing their own profiles, messaging, and meeting up with matches. They also

described their practices for evaluating other app user's profiles to filter out potential threats or 'red flags' (Albury et al., 2019).

The projects were informed by different disciplinary perspectives: Gillett's project took a criminological approach informed by the continuum of sexual violence (Kelly, 1988). Albury's project took a strengths-based health promotion approach informed by the disciplines of media and cultural studies. That is, it sought to identify and articulate app users' existing risk-mitigation strategies as opposed to assuming all app-users associated apps with harm or intrinsic vulnerability. While one project addressed the question of 'safety' via a criminological perspective, the other approached safety and risk through a lens of sexual health and emotional wellbeing. This comparative approach allowed us to make multiple insights into safety and surveillance. First, participants in both studies expanded the frameworks through which dating apps conceptualise safety, by speaking not only about sexual violence but about ordinary abuse, safer spaces and data security. Second, both projects offered insights into surveillance, from concerns over how apps were handling their data to the ways that users police one another based on their identity data. Third, the comparative approach allowed us to expose the way that safety responses and media reports are predominantly framed around heterosexual dating concerns and risks of sexual assault rather than more insidious forms of harm, including those facing LGBTIQ+ people relating to outing, harassment, transphobia, biphobia, whorephobia, sexual racism and entrapment.

User experiences of safety and harm in digital dating contexts

The purpose of Gillett's (2019) study was to investigate the range of men's behaviours women experience as intrusive on Tinder. Moving beyond a focus on criminal acts, Gillett (2019) drew on the continuum of sexual violence to account for 'ordinary', yet deeply insidious, intrusions that contribute to a broader culture of disrespect toward women. As an

app that facilitates online and in-person encounters, participants in the study described a range of intrusions both online and off. Most often, participants reported receiving intrusive messages through the platform's instant messaging service. The messages they disclosed often centred around attacks about their physical appearance and unsolicited requests for sex. Of the women who met men from Tinder offline, some participants reported being 'catfished'[1], stalked, and sexually or physically assaulted.

While participants described these experiences as intrusive and demonstrated the cumulative impact they had, they nevertheless conceptualised men's behaviour as typical and expected. One participant thought that men who sent harassing messages were engaging in "just typical guy stuff" (Becky). Even so, in digital dating contexts where the women often expected intimate intrusions, they shared the safety strategies they employed to protect themselves (see also Gillett, 2021). These practices included drawing on Tinder's safety features—such as the unmatch mechanism, which prevents communication between users—however, participants also relied on their own 'safety work' (Kelly, 2016) strategies including using platforms' technological features, turning off their mobile phones' geolocative capabilities and sharing their physical location with friends when they met with Tinder users on dates (see also Gillett, 2021). Participants also described the lengths they took to investigate their matches, such as by viewing their profiles on other platforms and asking them questions designed to interrogate their beliefs. Because the participants often thought that men were authentic on social networking websites tied to their own identities, they thought that men were more open about their motivations on social media platforms like Facebook. For this reason, there was

[1] Catfishing includes 'creating and portraying complex fictional identities through online profiles' (Nolan 2015, 53).

also a sense that digital dating contexts could be safer—in an intersubjective sense—than meeting men offline.

The challenge of prescribing a universal approach to 'safety' in app cultures was further reflected in the findings of Albury's study, which invited app users to define safety in their own terms. Responses addressed a range of topics including negotiating consent and safer sex practices; and dealing with racism, biphobia, transphobia, sexwork stigma, uninvited fetishisation and feelings of rejection on apps. Significantly, there was no consensus across all participants regarding the ideal app affordances in terms of 'safety features', given that app cultures varied widely across the range apps described.

For example, participants reported significantly different norms around disclosing sexual health status on 'men seeking men' apps (where HIV status was often openly disclosed in profiles) as opposed to 'straight' or 'women seeking women' apps, where the topic was generally avoided. There was general consensus, however, around the elements of profile design that signaled a potentially unsafe match. Profiles that did not include a picture of the user's face were mentioned in several workshop conversations as potential 'red flags' – signaling a potential 'cheater'. However, some users suggested that for men seeking men, 'headless torso' profiles might also signify the user was 'discrete', or did not identify as gay or bisexual – and this might be a valid safety strategy for members of communities that did not accept sex between men, and those living in regional or rural areas (Albury et. al. 2019). These findings suggest that even as dating apps are introducing a range of automated safety initiatives, it is not clear how well these new mechanisms address users' perceptions of safety – or the contextual factors that contribute to intersubjective safety for users of differing sexualities and genders. To the extent that abuse is context dependent (Dragiewicz et al.

2018), automated detection tools cannot accurately identify the range of content and behaviours that make people feel unsafe online.

"Let's add some emojis": Technological solutionism and app safety

As a platform that boasts being "the world's most popular app for meeting people" and "the highest grossing non-gaming app globally" (Tinder 2021), it is perhaps unsurprising that Tinder has historically focused on its business interests rather than safety initiatives. Participants in Gillett's (2019) study generally thought that Tinder's Reactions feature was an inadequate attempt to curb men's intrusive behaviour. One participant said: "Okay, instead of actually just making the usage of the app, you know, a little bit tighter, let's add some emojis" (Jade). Growing public, media, and state pressure on platforms to mitigate harm perpetrated through their networks, though, has seen companies take their users' safety more seriously. Tinder's House Rules and Safety Center outlines the platform's safety code of conduct and provides information on the app's safety and privacy features. However, many of Tinder's new safety features are reliant on automated decision-making systems to monitor users' behaviour. To mitigate 'catfishing' on the service, in 2020, the platform deployed a photo verification feature, which uses facial recognition technology to authenticate users' profile images (Tinder, n.d.). After piloting identity verification in Japan in 2019, Tinder has also announced plans to make the feature (ID Verification) available to all users over the coming year (Tinder, 2021a). On Tinder, these verification tools exist alongside automated content moderation systems, which use machine learning technology to detect harmful messages sent via the platform's private messaging service (Tinder 2021b). 'Are You Sure?', for instance, "uses AI to detect harmful language and proactively intervenes to warn the sender their message may be offensive" (Tinder, 2021b). Presumably, the same screening

technology underpins Tinder's 'Does This Bother You?' prompt, which asks users whether messages they have received are bothersome and provides a clear reporting avenue.

Automated or algorithmic moderation tools, which rely on either matching or predictive systems (Gorwa et al. 2020), detect what platforms understand as harmful or in violation of service rules. These tools may over detect overt 'offensive' language, while more ordinary intrusive content and behaviour—common in digital dating contexts (Gillett 2019)—may go unchecked. Perhaps most notably, Tinder's technological interventions operate on the logic that users do not deliberately conceal their physical appearance or send offensive messages on the service. Users who intentionally seek to deceive others about their physical appearance, for instance, can simply opt out of using the platform's photo verification feature. Users who are prompted by 'Are You Sure?' can choose to bypass the flagging system. This is a good thing for users whose messages are incorrectly flagged. But for conversations that users do experience as harmful, it is not clear what steps, if any, the platform takes to hold users to account for violating service rules.

We note, too, that automated identity-verification systems and platform 'real name' policies do not intrinsically increase safety for all app users – particularly if we take account of Möller and colleague's framework of intersubjective safety. Indeed, 'real' name policies, first introduced by Facebook under the guise of safety-through-transparency, are more accurately described as legal name policies, because pseudonyms or chosen names are not necessarily less 'real' or authentic for users. They have been criticized by queer communities, trans and gender diverse folk, sex workers, drag queens, Indigenous activists and multiple ethnically and linguistically diverse communities whose names are not deemed recognisable, and are argued to be "not about promoting safety but about rendering users transparent to markets and the state" (MacAulay and Moldes, 2016: 7).

‘Real’ name policies presume that users will be safer when their accounts are connected to their legal identities, because they conflate anonymity and pseudonymity with anti-social and unaccountable behaviour (van der Nagel and Frith 2015). However, as danah boyd points out, “many people are far LESS safe when they are identifiable” (boyd, 2011. n.p. emphasis in original). Activists being surveilled by police, survivors of intimate partner violence who may be stalked by ex-partners, trans folk who face legal barriers to changing their names and not want constant reminders of their deadnames, and sex workers who separate out their work and legal identities to avoid stigma or prosecution, all may be at risk when apps require them to use their legal names. In such cases, pseudonyms can be means of identity construction and compartmentalisation, used to avoid stigma, maintain boundaries, and to keep safe. Despite this, ‘real’ name policies and “default publicness” (Cho, 2018) form part of a broader trend towards authentication and verification found among social media platforms.

As Albury and colleagues (2021) have noted, app users with non-binary genders and sexualities (including genderfluid, bi and pansexual people) experience challenges on more ‘mainstream’ platforms (such as Tinder and Bumble) where filtering and matching menus only allow users to nominate the search categories ‘seeking men’ ‘seeking women’ or ‘seeking both’. A number of trans participants discussed their reluctance to share their real names – or images of their faces – on platforms where they felt heightened risk of transphobic reactions, including unwanted fetishization and threats of violence (Albury et al 2021). The need for safer spaces for queer, trans, gender non-conforming, two spirit, asexual and non-binary people has led to a diversification in the dating app market – for example, the dating app Lex prohibits cisgender men on their platform and supports text-only communication (López, 2019).

"What do they actually do with those reports?" User experiences of reporting and redress

For major global platforms that host millions of users, fostering safe digital dating contexts is no easy task. Participants in both Gillett and Albury's studies reflected on how they thought dating apps could make their services safer for users. Most broadly, Gillett's interview participants suggested that Tinder could update its technological interface to highlight the platform's reporting mechanism and safety information. Several of Gillett's interview participants, for instance, critiqued Tinder's reporting mechanism, which at that time, was obscured by three red dots in the corner of the instant messaging interface (see also Duguay et al., 2018; Gillett, 2021). While the placement and visibility of Tinder's reporting mechanism was important to these women, for those who did report rule violations, they also wanted the platform to transparently communicate the process and outcome of content moderation decisions.

Several participants in Gillett's study thought that Tinder's safety advice was manifestly inadequate. At the time of the interviews, Tinder listed a range of United States based hotlines and resources that users could engage with if they needed help. This led Samantha to question Tinder's motives for including safety advice at all: "even though they expand to other countries, they only really care about [the United States] because Americans are sue happy." As an app that operates in more than 190 countries, these women thought that Tinder should, at the very least, provide safety resources for users in each country the platform operates. Participants also thought that Tinder's safety tips reinforced harmful beliefs that women are responsible for navigating and preventing men's violence. Keira said: "Yeah, okay, right, so like, women have to regulate their behaviour, right." It is important to note

that Tinder has since updated its safety advice—developing a Safety Center—and changed its reporting mechanism.

Several participants in Gillett's study did not report abuse because they were unsure how the app would respond and what avenues the app could take to sanction the harm. Georgia wanted to know more about Tinder's decision-making process for user reports: "do they get one report and they get taken off or like what constitutes the threshold where they say to somebody "you can't use this app anymore"?" Specifically, Georgia wanted clarity on what would happen to the user who was reported, the processing of complaints, and the criteria and threshold that warranted different responses from the app, including account suspension, deletion, or police referral. She said:

> if people continuously report someone for saying something explicit, versus someone who was like "I've met up with this person and they've sexually assaulted me" like, how many messages does it take for somebody at head office to go "oh yeah okay, we're gonna delete this person"? Like yeah what's involved? The police? What do they actually do with those reports? (Georgia)

Other participants in Gillett's study did not report intrusions to Tinder because they understood the behaviour as expected and 'ordinary.' This led some participants to think that the app would not recognise users' behaviour as rule violations or take their reports seriously. Kimberly, for example, recalled: "I was worried that my interactions wouldn't be, I suppose, dramatic enough or bad enough…" Kimberly's fear that she was overreacting to intrusion was coupled with her understanding that men's behaviour was 'normal' and therefore did not justify reporting:

> It didn't come into my mind [the option to report Tinder users] and I think it's because we're just so used to […] it's part of the norm, so we just, sort of, accept it and just like well that guy's an idiot, onto the next one. So I really think that's why it didn't report him because I'm just like oh this is what guys do.

Participants who were willing to report abuse were disheartened by Tinder's response—or lack thereof. In both Gillett and Albury's studies, participants who did report violations of terms of service violations to the platform did not receive transparent information regarding the actions that platform had taken (or might take) to address the offending users' behaviour. One participant recalled reporting a user with whom she had the ability to match with on Tinder only days later. In the absence of transparent information regarding the outcomes and consequences of content moderation decisions, users question platforms' motives and actions (Suzor et al., 2019).

Additionally, some participants reported experiences common to victim-survivors of assault. Participants in Gillett's study avoided interpreting and labelling non-consensual sex as rape (see also Kelly 1988). One participant, Josie, described what she understood as the 'blurred line' between consensual sex and sexual assault:

> I was sitting there, like, did I just get abused? Did I just get assaulted? What happened? But I consented, but he was a little bit rough, and it made me very confused […] we have a very blurred line of what assault is because of the consent […] that line gets really blurred when you don't, you aren't sure of it.

Josie's response indicates that even criminal behaviour is normalised in broader heterosexual dating contexts, despite awareness raising campaigns (Aghtaie et al. 2018; Kelly 2012). Josie's account also points to an unintended outcome of the emphasis on consent, which she internalised as her responsibility to grant or refuse. She did not report the assault to Tinder or law enforcement because she did not know how to understand it herself. In other words, as Josie had consented to sex, her response indicates that no 'real rape' (Estrich 1987; Stanko 1985) was committed. These complex understandings suggest that law enforcement cannot rely on automated processes to understand the extent to which sexual violence is or can be facilitated by dating app affordances.

Albury's findings further suggest that there is no universal technological safety strategy that can meet the needs of all dating app users. For example, Max (23, trans-masculine) suggested that "if [app developers] are noticing a lot more trans people are starting to use the app, [they should consider] what can they do to make the trans people's experience on the app be better and safer." He then recounted a story of using Bumble during his transition, where changing his specified gender led to his account being re-set completely, deleting his existing matches and messages – an experience that was likely the result of a safety policy, designed to prevent users from falsifying their identity.

Many lesbian, bisexual and queer women expressed frustration that they were unable to limit men's access to their profiles on Tinder and Bumble, despite specifying that they were exclusively 'seeking women'. Blair (23, lesbian, trans female) was one of many women who reported this experience: "I'm not into men at all and the amount of apps that still show me men or say I've been liked by men, it's like why is my profile showing up to them all the time?" Agnes (22, pansexual female) emphasised this as the aspect of app use that concerned her the most, relating a friend's experience: "even when her profile was set to 'females only'

she was still swiping through those presenting as men and/or couples looking for a third. I experienced this heavily, too and it really turned me off using Tinder and Happn when exploring my sexuality." However, Ruby (29, bisexual female) used Tinder and Bumble in the context of a non-monogamous relationship with her male partner, and while she was sensitive to other women's frustration at being approached for threesomes, she believed "it's important to have … opportunities for non-mainstream relationships still in mainstream apps". Ruby suggested that couple's profiles could be marked differently to those of singles (including colour coding for different partners chatting in the same account) in order to assure "transparency" for app users.

While these findings are not representative of all user's experiences of apps, they suggest that app users are not universally opposed to technological approaches to user safety – including reporting features. However, these solutions are best focused on expanding app's transparency of communication, and capacity to recognise diversity in terms of users' genders and relationship styles and preferences.

Conflating surveillance with safety: Carceral solutionism as a response to harm

Some apps have adopted what might be termed carceral solutions to address technology-facilitated abuse, developing relationships with law enforcement or surveillance agencies. In April 2021, New South Wales Police announced they were in conversation with Tinder's parent company Match Group (which also owns OKCupid, Plenty of Fish and Hinge) regarding a proposal to gain access to a portal of sexual assaults reported on Tinder (McCormack, 2021; Gillett, Albury and Stardust, 2021). While Tinder later declined, the police also suggested using artificial intelligence to scan users' conversations for "red flags" (McCormack, 2021), with no further information on what kind of content, language or behaviour would be detected. This followed reports that, in the previous month, Match Group

had invested in the company Garbo to facilitate their users running 'background checks' against potential dates. With their name and phone number, users could be checked against public records and reports of violence, harassment or abuse, including information about their arrests, convictions and restraining orders. This announcement faced criticism that it would exacerbate ongoing criminal record discrimination, disproportionately impact communities who experienced excessive police targeting (including Black and Indigenous communities, people of colour and trans and gender diverse folk) and capture people who had criminal records for non-violent offences such as solicitation (Corrigan, 2021). The previous year, Match Group purchased a stake in the security app Noonlight, which invites users to log details about their dates, track their location, and allows users to hit a 'panic button' to alert law enforcement if they feel unsafe on a date. Although Match Group CEO stated that Noonlight would not share location data with the company (Blistein, 2020), an investigation by Gizmodo found that Noonlight had shared data with a range of third parties including Facebook and YouTube (Wodinsky, 2020).

These initiatives have been introduced in response to reports of serious violence by app users. Dating apps can be used to facilitate various offences including stalking, intimidation and theft. Digital media is increasingly used by domestic and family violence perpetrators for the purposes of control, coercion, and entrapment (Harris, 2020) as well as in sexual assault offences (Rowse, Bolt and Gaya, 2020). Dating app users may well participate in behaviour that constitutes a civil or criminal offence in their jurisdiction, and in such circumstances and others (such as applications for apprehended violence orders), chat and messaging histories may provide relevant evidence of the offence. Social media posts, saved chats and screenshots are regularly used in courtrooms, although each jurisdiction has different rules about the admissibility of digital evidence (Graves, Glisson and Choo, 2020). Apps store crucial information that can be valuable to law enforcement and

prosecutors, and dating apps have their own policies about how and when they provide user data to law enforcement.

However, providing specific evidence in response to subpoenas, court orders or search warrants is substantially different to affording police general access to chat histories and user reports. There are many reasons why users do not report abuse, either to apps or to police (Gillett, 2021). A wealth of evidence indicates that sexual violence survivors already have very poor experiences in trial proceedings, which are frequently influenced by rape myths in the courtroom (McDonald, 2020). Such proceedings are characterised by high attrition rates, low conviction rates and the prospect of being re-traumatised in court. Moreover, many marginalised communities such as sex workers report extremely poor experiences when reporting to police that place them at increased risk (Stardust, 2021). If dating apps make automated referrals to police or provide automatic access for law enforcement to their in-app reports, survivors are denied agency. For these reasons, we suggest apps seek to develop more tailored responses that do not involve automated police referral but instead seek to understand what survivors need in terms of redress, healing, and accountability.

Dating app data: Reasonable suspicion, predictive policing and police entrapment

For law enforcement agencies, dating app data provides a cost-effective form of out-sourced surveillance. Law enforcement agencies worldwide are pushing for various "backchannels" to access social media data without waiting for court approval – the FBI has proposed wiretapping encrypted communication services and the Canadian federal government has proposed requiring ISPs to disclose customer information to law enforcement without court orders (Trottier, 2012). In 2021, the Australian parliament passed the *Surveillance Legislation Amendment (Identify and Disrupt) Bill 2021*, which introduced what the non-

government organisation Digital Rights Watch refer to as a 'warrantless surveillance regime', permitting law enforcement agencies to covertly take control of social media accounts and impersonate users, add, copy, delete or alter data on devices, and overcome encryption to access entire networks where there is suspicion of serious online offences. These 'warrants' can be issued under an emergency authorisation without judicial approval (Digital Rights Watch, 2021).

Dating apps already offer law enforcement agencies an abundance of data that can be used for a broad range of purposes, from intelligence-gathering to entrapment. Such data can increase the likelihood of human rights abuses in jurisdictions where same-gender sexual activity is criminalised or stigmatised. For example, in her report *Digital Crime Scenes,* based on interviews with defense attorneys in Egypt, Lebanon and Tunisia, Afsaneh Rigot found that police are entrapping LGBTQ people on Grindr, Hornet and Kik and prosecuting them for 'moral crimes' and 'sexual acts against nature'. Rigot reports that prosecutors are increasingly relying on digital evidence from dating apps, including selfies, sexts, digital libraries and dating app chats (2020, 2022). In such instances, queer persecution is being "facilitated by increased availability of digital evidence and police searches of digital devices" (2022, 1). Similarly, in India, the prevalence of queerphobia on the part of police has been shown to facilitate hate crimes on gay dating apps (Sinha-Roy and Ball, 2021). In other contexts, law enforcement has also "used social media tools to monitor peaceful protests, assembled potentially innocuous social media activity as evidence for criminal conspiracy charges, [and] created fake profiles or impersonated individuals online" (Mateescu et al., 2015). There are cases of police performing unlawful searches of dating app users on their internal databases. For example, in 2020 a New South Wales Police Constable used his access to the Computerised Operational Policing System (COPS) database to look up

his Bumble date's police profile, viewing records linked to her residential address, former partner, apprehended violence order checks, transport offences and drug use (AAP, 2021).

There is also the risk that dating app data (including reports about user behaviour that do not meet the threshold of criminal conduct) could be used by law enforcement to develop targeted policing strategies for 'potential' offenders, contributing to net-widening. Law enforcement now increasingly relies on big data and machine-learning to make decisions throughout the criminal justice system, including about policing, bail, sentencing, risk, suspicion, and recidivism (Joh, 2016). And yet research demonstrates that automated hate-speech detection systems (such as those potentially employed by dating apps to flag harmful language) can harbour and perpetuate inherent racial biases (Davidson and Bhattacharya 2020; Noble 2018). Kate Crawford notes that machine learning systems are often trained on prejudicial data taken out of context, and that previous AI that purports to classify crimes has been built from inaccurate and skewed police training data (Crawford, 2021: 95). The enthusiasm of law enforcement to access dating app data must be understood in this context.

The risks of predictive policing are of particular concern given the lack of transparency and privatised contexts in which such datasets are created. As Sarah Brayne points out, based on her empirical interviews and observations with the Los Angeles Police Department, although states have always collected data on their citizens, "[w]hat is new and important is that the state is relying more heavily on private vendors and platforms to collect, store, share and analyze data about its citizenry" (Brayne, 2021: 5). The algorithms used by dating apps such as Tinder to detect 'red flags' monitor chats, identify harmful language and sort content are proprietary and protected as trade secrets or by non-disclosure agreements. This means that there may be even less transparency into how law enforcement agencies make decisions,

exercise discretion and can be held into account, because we cannot openly evaluate whether such suspicion or discretion is reasonably exercised.

Click 'yes' to sex: Documenting and datafying consent

If dating apps alone are not currently able to ensure user safety via reporting mechanisms, some police departments have fixated on datafying consent through third-party consent apps (or consent plug-ins) as a solution to gendered harm. In March 2021, the NSW Police Commissioner Mick Fuller proposed the introduction of a 'consent app', suggesting that prospective sexual partners could record their consent on an app prior to having sex (Fuller, 2021). Likening it to the Australian government's COVIDSafe app, whereby users check-in at venues to track their movement in order to facilitate contact-tracing by health departments, Fuller suggested that an app might "protect people" who are dating and work to "[keep] people out of the justice system" (McGowan, 2021). He proposed that users would enter their name and age, declare that they understand sexual consent and then 'accept' or 'deny' before proceeding to sexual activity. The app was inspired by the iConsent app launched in Denmark in December 2020, whose logo depicts a handshake in the shape of a heart. In its product information, iConsent claims that users receive "the opportunity to document your consent to intercourse" in less than 30 seconds (iConsent, 2021). The data is encrypted and stored on the app for potential use later in criminal investigations.

Consent apps such as these, which approach consent as a static and irrevocable transaction, do not take account of the fact that consent is dynamic and can be withdrawn. Criminal law reform is now moving towards new frameworks that understand consent as 'communicative', a mutual and ongoing interaction that involves seeking feedback and taking reasonable steps to inquire about another person's state of mind. In 2020, the NSW Law Reform Commission recommended that the law ought to include a new subdivision on communicative consent,

clarifying that consent "is a continuous process of mutual decision-making" (NSWLRC, 2020: 47). They further state that "consent once given may be withdrawn at any time or its scope altered. Consent to one kind of sexual activity does not imply consent to any other activity" (p.51). Ongoing law reform in multiple jurisdictions makes evident that parties do not simply consent to sexual activity in general; rather, consent is specific. Parties may consent to particular types of sexual activity, with particular people, at particular times, on particular conditions, and always have a right to change their mind.

It is easy to see how consent apps could be used in sexual assault trials by defence counsel in attempts to establish that survivors had consented to sexual activity. The law on sexual offences in some jurisdictions has now evolved to accept that sexual assault can occur during marriage, that sex workers can withdraw consent, and that victims can be so influenced by alcohol or drugs that they do not have capacity to consent (Graycar and Morgan, 2002). Despite this, myths about consent persist, particularly during cross-examination (Quilter, 2021). The international #metoo movement has drawn attention to the way rape myths permeate both the courtroom and broader society, to insinuate that survivors are lying, were asking for it, or do not appear traumatised enough (Fileborn and Loney-Howes, 2019).

A 'yes' click to sexual activity on a consent app cannot prove that a person was consenting at the time of sexual activity or to the particular act in question, because no person is contractually obligated to follow through with sex, regardless of whether they express interest or even agree at some point. In both Denmark and Australia, proposals to introduce consent apps have met with criticism (McGowan, 2021). The Australian app was denounced by feminist scholars, public commentators and parliamentarians invariably as misguided and ill-conceived (Nguyen and Cockburn, 2021; Mazzoni, 2021). The Police Commissioner was criticised for failing to listen to the many evidence-based proposals for addressing sexual

violence and for presenting a naïve solution that was more likely to protect perpetrators than survivors of sexual assault (Bath, 2021).

Role-modelling consent: Personal information, data sharing and app transparency

While stand-alone sexual consent apps are unlikely to increase app users' safety in intersubjective terms, dating apps can learn from discourse on sexual consent, communication and transparency by role-modelling these practices on the apps themselves, especially in relation to their transparency about how they share user data. Some research has explicitly examined the privacy vulnerabilities of dating apps to find that many contain detailed personal information (from stored messages and location readings to sending unencrypted private conversations, profile pictures and private images) that could be used by forensic practitioners seeking to find evidence of criminal activity (Farnden, Martini, and Choo 2015). The kinds of personal information that social media companies provide to law enforcement ranges from contact information and messages to photographs the user has uploaded or been tagged in, in addition to affiliations, groups and friends lists (Trottier, 2012). Even apps that claim not to store user content may still store metadata, and forensic software can be used to recover photos and chats.

While dating apps have faced relative impunity for poor data-sharing practices, some regulators are beginning to take this issue more seriously. For example, the European Union's General Data Protection Regulations 2016/679 (GDPR) seek to give individuals greater control over their personal data by regulating how it is transferred. In particular, Article 9 of the GDPR prohibits the processing of data concerning a person's sex life or sexual orientation. This provision would prevent dating apps from processing user data where an individual has listed their sexuality or sexual preferences in their dating app profile.

In January 2021, the Norwegian Data Protection Authority published an intention to issue a 100million Norwegian Krone administrative fine to the dating app Grindr, for not complying with the GDPR on consent. In their preliminary conclusion, they found that Grindr had unlawfully shared user data to a number of third parties for advertising purposes, including GPS location and user profile data. Because Grindr markets itself as a dating app for gay and bisexual men, simply disclosing that a person is using Grindr provides information about their sexual orientation. The Norwegian Data Protection Authority regarded this as a serious case given that users were not able to exercise real or effective control over the sharing of their data, were not properly informed about what they were consenting to and were pressured into giving consent through Grindr's business model (Datatilsynet, 2021). In particular, the NDPA considered that consent is required for "intrusive profiling and tracking practices for marketing or advertising purposes". Grindr's fine was not insignificant – the Guardian reports that it amounted to 10% of their global annual revenue (Hern, 2021). The case was prompted by an investigation by Norway's Consumer Council into the data practices of 10 apps, which found that Grindr stood out for the lack of information contained in its privacy policy (Hern, 2021).

If this case marks the beginning in a trend towards accountability of dating apps in terms of how they handle their user's personal data, then dating apps ought to be especially transparent about how they may use, sell, or share user data and far more intentional about how they obtain informed and specific consent from their users to do so. Research among people with HIV, trans and gender diverse people, sex workers, and gay and bisexual men in Australia found that those populations had very low levels of trust in digital health that was influenced by relational and structural factors more so than technical design (Newman et al, 2020). If apps are committed to building cultures of consent, and want to engender the trust of their users, they ought to begin by dramatically improving their own consent practices

regarding the sharing of personal and sensitive data. This includes introducing privacy policies that offers users greater transparency and decision-making over what happens to their data alongside accessible information and options for how their reports of harmful conduct are considered and actioned.

Afsaneh Rigot makes a number of recommendations for dating apps, such as the ability send timed/ephemeral messages, the use of double security PINs and secret folders to restrict access to police, panic and self-destruct buttons in the face of interrogation or entrapment, warnings to notify users when another user takes a screenshot, not linking usernames to registered phone numbers, not saving photos directly into a gallery, and supporting local legal aid funds (2022, 138-144). In doing so, Rigot calls for "design from the margins", a methodology centred on justice, human rights and the experiences of those most impacted by technology: "By understanding who is most impacted by social, political and legal frameworks, we can also understand who would be most likely to be a victim of the weaponization of certain technologies" (2022, 4-5).

Conclusions: Building consent culture

On a structural level, the enthusiasm of some apps to foster relationships with law enforcement or introduce new surveillance software is fraught. The reliance on surveillance and law enforcement as the primary response to interpersonal harms – whether they occur in-app or are facilitated via the app – misguidedly assumes that incarcerating people will change consent cultures or build more equitable sexual societies. It is clear that dating app users are not 'at risk' on apps, but 'safe' in other (off-line) contexts. Rather, dating app affordances and user cultures have reinscribed (and in some cases amplified) structural inequalities facing women, trans and gender diverse people, people of colour, people with disabilities, and other marginalised and stigmatised groups.

Returning then to the concept of intersubjective safety, whose sexualities and subjectivities are currently considered to be 'safe'? For decades, Black communities, Indigenous communities, and people of colour have critiqued the carceral politics of white feminists, which positions police as saviours and the criminal legal system as the answer to a suite of social, cultural, and economic issues (Phipps, 2020). In doing so, white feminists have contributed to a mass incarceration crisis and agitated to expand the powers of police and prosecutors whilst serving to reinforce stereotypes of white victimhood (Gruber, 2020). Meanwhile, turning to law enforcement as a solution for sexual violence does not support survivors or offenders. Prisons frequently produce conditions where patterns of abuse and criminality repeat, rather than fostering cultures of consent or transformative justice that are necessary to change societal conditions or prevent future sexual violence (Levine and Meiners, 2020). If dating apps are invested in building consent culture and preventing technologically-facilitated abuse, it will require a different approach to technological or carceral solutionism.

Major platforms across the globe are now grappling with how to moderate enormous volumes of content (Gillespie 2020). But for many platforms—dating and social media more broadly—their approach to safety has been to add 'safety solutions' and stir. This approach has seen platforms focus on moderating discrete incidents, perhaps reflecting "an extension of Western criminal justice systems that prioritize retribution over structural change via accountability and repair" (Schoenebeck and Blackwell 2021). More research is required to investigate the partnerships and relationships between big technology companies and state law enforcement agencies, including the extent of their data sharing and procurement practices. Indeed, as Tinder and other dating apps turn to employ third party services, and establish law enforcement partnerships, there is a good argument that these initiatives weaponise women's safety to justify police and state surveillance tactics, expansion of police

power and, inevitably, more punitive carceral measures. The private, proprietary nature of the software used to identify and assess risky, unsafe or harmful behaviour does not promote trust among dating app users that the discretionary decisions made by apps or police are reasonable, just or even lawful. But "we cannot end violence by doing violence" (Phipps 2020, 168). Pursuing a techno-carceral approach that equates surveillance with safety, technology with progress and police with justice will only serve to deepen and encode existing inequalities on dating apps.

Datafying consent will not protect dating app users. However, understandings of sexual consent as a dynamic, interactive and communicative practice can help shape dating apps' policies towards safety and data privacy. A genuine investment in building consent culture requires multiple components, including education, resourcing, and building the capacity of users to negotiate mutual pleasures and boundaries, in addition to centering the experiences of survivors by offering material support and transparent options that respect their agency and decision-making. In this article, we have drawn on empirical accounts of app use – and popular media reporting – to problematise commonsense assumptions about dating apps, safety, technology, policing and surveillance. A critical criminological perspective offers us a useful lens to think about accountability by transforming the systems under which consent is navigated, while a public health approach demonstrates the value of adopting a nuanced and contextual approach to gender and sexual diversity. Given that dating apps collect substantial amounts of intimate data about their users, they have a unique opportunity to lead by example. If dating apps are committed to advancing consent culture, and not simply to quick reputational fixes, they could actively build in avenues for users to expressly consent to (and withdraw from) specific uses of their data. This includes refusing intimate data from being sold, monetized or shared with law enforcement.

Conflicts of Interest

The Authors declare that there is no conflict of interest.

Funding

The authors disclosed receipt of the following financial support for the authorship and publication of this article: This work was supported by the Australian Research Council Centre of Excellence for Automated Decision-Making and Society [grant number CE200100005] and the Australian Research Centre Linkage Project LP160101687.

References

Aghtaie, N, Larkins C, Barter C, Stanley N, Wood M and Øverlien C (2018) Interpersonal violence and abuse in young people's relationships in five European countries: Online and Offline normalisation of heteronormativity. *Journal of Gender-Based Violence* 2(2): 293-310.

Albury, K, Burgess J, Light B, Race K, and Wilken R. (2017). Data cultures of mobile dating and hook-up apps: Emerging issues for critical social science research. *Big Data & Society*, *4*(2), 2053951717720950.

Albury, K, Byron P, McCosker A, Pym T, Walshe J, Race K, Salon D, Wark T, Botfield J, Reeders D, and Dietzel C. (2019). Safety, risk and wellbeing on dating apps. Final Report. Swinburne University of Technology.

Albury, K., McCosker, A., Pym, T., & Byron, P. (2020). Dating apps as public health 'problems': cautionary tales and vernacular pedagogies in news media. *Health Sociology Review*, *29*(3), 232-248.

Albury, K., Dietzel, C., Pym, T., Vivienne, S., & Cook, T. (2021). Not your unicorn: Trans dating app users' negotiations of personal safety and sexual health. Health Sociology Review, 30(1), 72-86.

Blistein J (2020) Tinder app to add panic button for extremely bad dates. *Rolling Stone*. 23 January. Available at: https://www.rollingstone.com/culture/culture-news/tinder-safety-features-panic-button-alarm-noonlight-941838/ (accessed 14 October 2021).

Blunt, D, and Stardust, Z (2021) Automating Whorephobia: sex, technology and the violence of deplatforming: An interview with Hacking//Hustling. *Porn Studies*, *8*(4), pp. 350-366.

boyd d (2012) Privacy and security: The politics of 'real names'. *Communications of the ACM*. 55(8): 29-31.

boyd d (2011) 'Real Names' Policies are an Abuse of Power. Apophenia. Available at https://www.zephoria.org/thoughts/archives/2011/08/04/real-names.html (accessed 10 March 2022).

Brayne S (2020) *Predict and surveil: Data, discretion, and the future of policing*. USA: Oxford University Press.

Brown I (2015) Social media surveillance. The *I*nternational *E*ncyclopedia of *D*igital *C*ommunication and *S*ociety. 1-7.

Cho A (2018) Default publicness: Queer youth of color, social media, and being outed by the machine. *New Media & Society*, 20(9), 3183-3200.

Corrigan H (2021) Tinder's plan for criminal record checks raises fears of 'lifelong punishment'. *The Guardian*. 13 April. Available at:

https://www.theguardian.com/technology/2021/apr/13/tinder-plan-criminal-record-checks (accessed 14 October 2021).

Crawford K (2021) *The Atlas of AI*. United States: Yale University Press.

Cross C, Parker M and Sansom D (2019) Media discourses surrounding 'non-ideal'victims: The case of the Ashley Madison data breach. *International Review of Victimology* 25(1): 53-69.

Davidson T and Bhattacharya D (2020) Examining racial bias in an online abuse corpus with structural topic modeling. *arXiv preprint arXiv:2005.13041*.

Datatilsynet (2021) Intention to issue €10 million fine to Grindr LLC. *Datailsynet.* 26 January. Available at: https://www.datatilsynet.no/en/news/2021/intention-to-issue--10-million-fine-to-grindr-llc2/ (accessed on 14 October 2021).

Digital Rights Watch (2021). Australia's new mass surveillance mandate. 2 September. Available at https://digitalrightswatch.org.au/2021/09/02/australias-new-mass-surveillance-mandate/ (accessed 14 October 2021).

Dragiewicz M, Burgess J, Matamoros-Fernández A, Salter M, Suzor NP, Woodlock D and Harris B (2018). Technology Facilitated Coercive Control: Domestic Violence and the Competing Roles of Digital Media Platforms. *Feminist Media Studies* 18(4): 609-625.

Dragiewicz, M and DeKeseredy WS (2012). Claims about women's use of nonfatal force in intimate relationships: A contextual review of Canadian research. *Violence Against Women* 18 (9): 1008-1026.

Dubrofsky RE and Magnet S (2015). *Feminist Surveillance Studies*. Durham, NC: Duke University Press.

Duguay, S, Burgess J and Suzor NP (2018). Queer women's experiences of patchwork platform governance on Tinder, Instagram, and Vine. *Convergence: The International Journal of Research into New Media Technologies*. 26(2): 237-252.

Estrich, S (1987). *Real Rape*. Cambridge: Harvard University Press.

Fileborn B and Loney-Howes R (2019) *#MeToo and the politics of social change*. Palgrave Macmillan.

Farnden J, Martini B and Choo KKR (2015) Privacy risks in mobile dating apps. Proceedings of 21st Americas Conference on Information Systems. 13-15 August. Association for Information Systems.

Gillespie T (2020) Content Moderation, AI, and the Question of Scale. *Big Data & Society*, Epub ahead of print August 21. DOI: 10.1177/2053951720943234.

Gillett, R (2019) *Everyday violence: Women's experiences of intimate intrusions on Tinder.* PhD thesis, Queensland University of Technology.

Gillett, R (2021) 'This is not a nice safe space': investigating women's safety work on Tinder. *Feminist Media Studies*.

Gillett, R, Albury, K, and Stardust, Z (2021) NSW Police want access to Tinder's sexual assault data. Cybersafety experts explain why it's a date with disaster. *The Conversation*, 28 April 2021.

Gorwa R, Binns R and Katzenbach C (2020) Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance. *Big Data & Society* 7, Epub ahead of print February 28. DOI: 10.1177/2053951719897945.

Graves L, Glisson WB and Choo KKR. (2020) LinkedLegal: Investigating social media as evidence in courtrooms. *Computer Law & Security Review* 38:105408.

Graycar R and Morgan J (2002) *The Hidden Gender of Law*. Second edition. Sydney: Feminist Press.

Green J (2019) Who's Watching You? More Than Just Your Date: Surveillance and Privacy on Grindr. *On Politics: Undergraduate Journal of Political Science*. 13 (1): 58-74.

Harris B (2020) Technology and violence against women. In: Walklate S, Fitz-Gibbon K., McCulloch J and Maher JM (Eds) *The Emerald Handbook of Feminism, Criminology and Social Change.* United Kingdom: Emerald, pp. 317-336.

Joh EE (2016). The new surveillance discretion: automated suspicion, big data, and policing. *Harvard Law and Policy Review*. 10(1): 15-42.

Kelly L (1988). *Surviving sexual violence*. Minneapolis: University of Minnesota Press.

Kornstein H (2019). Under her eye: Digital drag as obfuscation and countersurveillance. *Surveillance & Society* 17(5): 681-698.

Light, B, Burgess J and Duguay S (2016) The walkthrough method: An approach to the study of apps. *New Media and Society* 20(3): 881-900.

López, C (2019). How to use Lex, a queer dating app with no profile pictures or cisgender men. *The Insider.* November 21.

MacAulay M and Moldes M.D (2016) Queen don't compute: reading and casting shade on Facebook's real names policy. *Critical Studies in Media Communication* 33(1): 6-22.

Mateescu A, Brunton D, Rosenblat A., Patton D, Gold Z, boyd d (2015). Social Media Surveillance and Law Enforcement. In: Data & Civil Rights: A New Era of Policing and Justice. October 27. Available at https://datasociety.net/wp-content/uploads/2015/10/Social_Media_Surveillance_and_Law_Enforcement.pdf (accessed 14 October 2021).

McCormack A (2021) Tinder 'in conversation' with police departments, as NSW Police propose new dating app safety measures. *Triple J, Hack.* Australian Broadcasting Service. 26 April. Available at: https://www.abc.net.au/triplej/programs/hack/tinder-announces-new-safety-measures-artificial-intelligence/13317896 (accessed 14 October 2021).

McDonald E (2020) Rape myths as barriers to fair trial process: comparing adult rape trials with those in the Aotearoa Sexual Violence Court Pilot. Christchurch: University of Cantebury Press.

Naidoo, NA (2021) We're partnering with Bumble to bring Bloom to their users. Chayn. August 5. Available at https://blog.chayn.co/were-partnering-with-bumble-to-bring-bloom-to-their-users-97128ce28e7f (accessed 14 October 2021).

Newman C, MacGibbon J, Smith AKJ, Broady T, Lupton D, Davis M, Bear B, Bath N, Comensoli D, Cook T, Duck-Chong E, Ellard J, Kim J, Rule J and Holt M. (2020) Understanding trust in digital health among communities affected by BBVs and STIs in Australia. Sydney: UNSW Centre for Social Research in Health.

Nolan MP (2015) Learning to circumvent the limitations of the written-self: The rhetorical benefits of poetic fragmentation and internet 'catfishing.' *Persona Studies* 1(1): 53-64.

Phipps A (2020) *Me, not you: The trouble with mainstream feminism,* Manchester: Manchester University Press.

Rigot A (2020) Egypt's Dangerous New Strategy for Criminalizing Queerness. *Slate*. 30 December. Available at https://slate.com/technology/2020/12/egypt-lgbtq-crime-economic-courts.html (accessed 14 October 2021).

Rigot A (2022) Digital Crime Scenes: The Role of Digital Evidence in the Persecution of LGBTQ People in Egypt, Lebanon and Tunisia. *Article 19*. Available at https://www.article19.org/wp-content/uploads/2022/03/Digital-Crime-Scenes-Report-3.pdf (accessed 10 March 2022).

Rowse J, Bolt C and Gaya S (2020) Swipe right: the emergence of dating-app facilitated sexual assault. A descriptive retrospective audit of forensic examination caseload in an Australian metropolitan service. *Forensic Science Medicine and Pathology* 16(3): 71–77.

Schoenebeck A and Blackwell L (2021) Reimagining Social Media Governance: Harm, Accountability, and Repair. *SSRN Electronic Journal*. Epub ahead of print. 29 July. DOI: 10.2139/ssrn.3895779.

Sinha-Roy, R. and Ball, M., 2021. Gay Dating Platforms, Crimes, and Harms in India: New Directions for Research and Theory. Women & Criminal Justice, pp.1-17.

Stanko, E (1985) *Intimate intrusions: Women's experiences of male violence.* London: Routledge.

Stardust, Z, Treloar, C, Cama, E, and Kim, J (2021) 'I Wouldn't Call the Cops if I was Being Bashed to Death': Sex Work, Whore Stigma and the Criminal Legal System. *International Journal for Crime, Justice and Social Democracy*, *10*(2).

Stardust Z and Caldwell H (2022) 'Archetypal Sluts: Payment of sex workers as a condition of consent' in Yvette Russell and Kate Gleeson, *New Directions in Sexual Violence Research*, Routledge, forthcoming.

Stoicescu M and C Rughiniş (2021) Perils of digital intimacy. A classification framework for privacy, security, and safety risks on dating apps.23rd International Conference on Control Systems and Computer Science (CSCS): 457-462. DOI:10.1109/CSCS52396.2021.00081.

Suzor, NP, Myers West S, Quodling A and York J (2019) What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation. *International Journal of Communication* 13: 1526–1543.

Tinder (2017) Tinder Launches Reactions. Available at https://www.tinderpressroom.com/tinder-launches-reactions (accessed 14 October 2021).

Tinder (n.d) "What is Photo Verification?" Available at https://www.help.tinder.com/hc/en-us/articles/360034941812-What-is-Photo-Verification- (accessed 14 October 2021).

Tinder (2021a). Tinder Will Add ID Verification Option Following the Successful Rollout of Photo Verification. Available at https://www.tinderpressroom.com/2021-08-16-Tinder-Commits-to-ID-Verification-for-Members-Globally,-a-First-in-the-Dating-Category (accessed 14 October 2021).

Tinder (2021b). Tinder Introduces Are You Sure?, an Industry-First Feature That is Stopping Harassment Before It Starts. Available at https://www.tinderpressroom.com/2021-05-20-Tinder-Introduces-Are-You-Sure-,-an-Industry-First-Feature-That-is-Stopping-Harassment-Before-It-Starts (accessed on 14 October 2021)/

Trottier D (2012) Policing social media. *Canadian Review of Sociology* 49(4): 411-425.

Van der Nagel, E. and Frith J (2015) Anonymity, pseudonymity, and the agency of online identity: Examining the social practices of r/Gonewild. First Monday 20(3).

Walsh JP and O'Connor C (2019) Social media and policing: A review of recent research. *Sociology compass* 13(1) e12648.

Wark, M (2019) *Capital is dead: is this something worse?* London: Verso

Wodinsky, W (2020) Tinder's new panic button is sharing your data with ad-tech companies. *Gizmodo*. 24 January. Available at https://gizmodo.com/tinders-new-panic-button-is-sharing-your-data-with-ad-t-1841184919 (accessed on 14 October 2021).