palgrave pivot

Gaming the Dynamics of Online Harassment

Kevin Veale

palgrave macmillan

Gaming the Dynamics of Online Harassment

"This book thought-provokingly details overlaps between the operations of alternate reality games and online harassment communities. Framing the "rabbit hole" as a starting point, Veale deftly uses this comparison to outline how participation in both types of community can develop from simple curiosity to intense involvement in a like-minded community with a shared goal. Veale also wisely recognises that hate online is a reflection of broader social issues and dynamics in a way that highlights both potential solutions and the significant challenges of online harassment."

—Dr Michael S. Daubs, Programme Director and Senior Lecturer of Media Studies, *Victoria University of Wellington, Aotearoa-New Zealand*

"This book is a fresh and original take on the topic of online hate and harassment. Veale provides an insightful and comprehensive account of the major developments and milestones in the history of online harassment campaigns, explaining their complex dynamics, both at the level of personal interaction and platform complicity. This book is a timely contribution to a steadily growing body of work that seeks to reconceptualise online hate as terrorism, and—most importantly—to find solutions to this problem."

—Dr Debbie Ging, Associate Professor of Media Studies, *Dublin City University, UK*

Kevin Veale

Gaming the Dynamics of Online Harassment

palgrave
macmillan

Kevin Veale
School of Humanities, Media and Creative Communication
Massey University
Wellington, New Zealand

ISBN 978-3-030-60409-7 ISBN 978-3-030-60410-3 (eBook)
https://doi.org/10.1007/978-3-030-60410-3

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

I have been lucky enough for online spaces to feature significantly in my life: some of my oldest friends are people I met online on a videogame forum. The fact that the internet is monstrous to people is something I have grown up with, and it has always been intuitively obvious that it focuses its monstrousness on women and people who do not fit the white, straight, cisgender norm.

This project began as an outgrowth of those experiences: my preferred field of study is exploring storytelling and the way storytelling experiences are changed by the media forms they are communicated through. Alternate reality games (ARGs) are one of the arenas of storytelling I have explored, particularly the ways that they are impossible to predict ahead of time, deploy terrifying ability to solve problems and think laterally, and often appear in areas people are not expecting them to be.

Spending life online and digging into how ARGs work means that when online culture began to change in the 2000s and 2010s, it looked familiar to me. I wrote about how arguably the online culture of 4Chan self-organised into patterns which *looked* almost like ARGs. In 2012, I saw the campaign against Anita Sarkeesian's *Tropes vs Women in Videogames* series stumble into becoming an ARG focused on her destruction. In 2014, I saw Gamergate turn itself into a malevolent ARG on purpose, with practice, because many of the same people from the campaign against Sarkeesian were gleefully involved.

And it never stopped.

I and my whole cohort have lived through nearly a decade under the shadow of organised, structured campaigns to destroy people we know

and care for, campaigns which preferentially target the most vulnerable among us and which seek to terrorise us all into silence.

This book exists because I started asking whether the fact that online harassment works like ARGs could be used against them, and that question took me down a rabbit hole about how the status quo works and what it would take to change things.

Change is possible, but there are plenty of people who will resist it at a large number of levels, and this book explores how, why and, with any luck, how we can try to circumvent them.

Wellington, New Zealand Kevin Veale

Acknowledgements

This research project was supported by the School of Humanities, Media and Creative Communication within the College of Humanities and Social Sciences at Massey University in Aotearoa-New Zealand through the Massey University Research Fund. Thanks to everybody in the School who helped facilitate the development and publication of this book: your help and support is always greatly appreciated.

A vast number of people have contributed to this project across the years I have been quietly working on it; many of whom have actively asked not to be credited in a book about online harassment because visibility would put them at risk. This broad category includes the following:

- My wonderful spouse, who has always done a huge amount of the daily labour that is often unappreciated and invisible in academia involved in keeping someone like me fed, watered, clothed and happy, and in a position to do work like this.
- My editors at Palgrave Macmillan, who fielded a wide variety of odd questions and offered excellent support and guidance at every stage of the project, from the proposal through to peer-review and publication. The comments from the anonymous reviewer were also excellent, constructive and greatly appreciated.
- Colleagues at my current university who provided substantial support and mentorship across the project at many levels and in many ways.

- Media critics and scholars discussing videogames and/or online harassment, some of whom have left a combination of the internet or the field as a result of harassment.
- A large number of people working online and in online spaces who have made concrete suggestions for those spaces could be improved and made safer.

I also want to mention Alex 'Kazz' McDougall, a friend of many years who lost his life in 2018 partly as the result of a sustained campaign of online harassment. He deserved to be treated better by both the world and the online spaces within it that he loved so well, and hopefully, this book provides some tools for making them less unsafe for someone else in future.

Contents

CHAPTER 1

Introduction: The Breadth of Harassment Culture and Contextualising Gamergate

Well lads, it's time to stop shitposting and time to make a real-life effort post.
—Macklin (2019); Rowe (2019)

This statement comes from a post on the white-supremacist and child-porn haven, 8chan, before the terrorist attack on the 15th of March 2019 targeting the Al-Noor Mosque and Linwood Avenue Islamic Centre in Christchurch, Aotearoa-New Zealand.[1] The terror attack was incubated in online spaces and spoke back to them, inspiring a series of further terror attacks around the world which made references to it in different ways, some oblique and some very directly. The consequences of online abuse and harassment have never *stayed* online. The Christchurch attacks are one of the most visible tips of an iceberg which has existed for many years, despite being seen as just a 'virtual' problem and not being taken seriously by law enforcement and legal systems around the world (Citron 2014, 100–102; DePass 2016; Elwell 2014; Golding and Van Deventer 2016, 51, 98; Lehdonverta 2010; Phillips 2015a, 41; Shaw 2014; Shepherd et al. 2015). It is tempting to conclude that the Christchurch attacks and other online-originated harassment campaigns[2] that extend fingers into the real world are a new development. However, they have a long history, and although we have definitely seen an increase in the amount of abuse

K. Veale, *Gaming the Dynamics of Online Harassment*,
https://doi.org/10.1007/978-3-030-60410-3_1

online, the escalations in organisation applied to delivering that abuse are more impactful than the rise in volume in itself.

This chapter will contextualise the history of online harassment, exploring the ways that it is part of broader cultures of harassment which happen to play out in online spaces, and will discuss case studies from the ongoing harassment campaigns which became both more organised and more visible over the last decade.

Contextualising Online Abuse and the Breadth of Harassment Culture

Abuse and harassment are so ubiquitous online that for decades anyone who does not like it has been told to toughen up or get off the internet (Citron 2014, 79–80). Kathy Sierra was systematically terrorised in 2004, and after she discussed her harassment, a neo-Nazi serial-harasser posted her social security number and home address online (Citron 2014, 35–39; Sierra 2014). In 2006, after the suicide of 13-year-old Mitchell Henderson, 4chan unleashed coordinated harassment on his family because the online community collectively known as Anonymous found the circumstances surrounding the death amusing—with at least one member of Anonymous going so far as to physically visit Mitchell's grave (Phillips 2015a, 28–29). Since 2010, online memorials for people who have died have been lightning rods for abuse (Phillips 2010, 2015a, 71–94). The possibility of being targeted by forces seeking to reach out and damage your ability to live your life has been part of the background-radiation of existing online for decades—although it is not background-radiation that affects everyone equally.

Cultures of abuse online disproportionately target anyone identifiable as not being a white, straight, cisgender man (Banet-Weiser 2015; Citron 2014, 13–15; Condis 2015, 2018; Massanari 2015, 2018; Phillips 2015a, 42, 53, 164; 2015b). The cultural bones of the internet are built up of technolibertarian/utopian ideals of freedom which argue everyone is equal (and equally 'disembodied') in online spaces,[3] and this flows into gaming contexts as well (Condis 2015, 201, 206–7; Lindbergh 2020; Massanari 2015, 5; Shepherd et al. 2015, 4; Turner 2006). Rubin et al. argue that this framework allows social media platforms and technology companies to hide behind a 'façade of neutrality' that ignores disparate impact on marginalised groups (Rubin et al. 2020, 1). These assumptions

imply that since everyone is equal, anyone drawing attention to the fact that they are not a white, straight, cisgender man is inauthentic, since they are 'choosing' to be more invested in that facet of their identity than in being a gamer or a citizen of the web (Citron 2014, 78). Megan Condis highlights the tensions produced by these assumptions in exploring conflict tied to the censorship of terms like 'gay' and 'lesbian' on the forums for *Star Wars: The Old Republic* (SW:TOR).

> TOR fans argued over whether an online game is an appropriate venue to discuss the sexual politics and the problem of heteronormativity in virtual worlds. What was often framed by the participants as a benevolent desire to prevent political and ideological conflict from leaking into gaming and ruining its unique attractions manifested as the maintenance of a heterocentric power structure. True gamers and fans are assumed to be straight (or, if they are queer, it is assumed that they will remain in the closet while participating in the gaming forum), and out queer gamers and their allies are flagged as disruptive and harmful interlopers. This stance implies that BioWare would be doing its real fans (the ones they rely on to sustain their profit margins) a disservice were it to cater to the desires of queer players by making the forum community queer friendly. A similar debate arose 2 years later when BioWare made the decision to include gay male romance options in their popular single player role-playing game franchise, *Dragon Age*. (Condis 2015, 199–200)

As Condis identifies, one of the assumptions tied to the theoretical equality of online and gaming spaces is that they are 'apolitical,' meaning that anyone seeking to change the representational dynamics in those spaces is 'bringing in politics' to an otherwise politically neutral environment.[4] These assumptions lead to the conclusion that anyone seeking change is not authentic and is dragging unwanted politics in from 'outside' (Beauchamp and Condis 2019; Condis 2019)—and must be resisted, particularly when those conversations seem to be resulting in some changes to the culture (Golding and Van Deventer 2016, 130–31).

Any change from the status quo represents a threat, regardless of its context, scope or scale. If it comes from 'outside,' it must be resisted (Beauchamp and Condis 2019; Condis 2019). Part of the reasoning behind this involves an understanding that culture is a 'zero-sum' game where no compromise is possible and where progress in one area requires defeats in another (Shepherd et al. 2015, 4–5). Mia Consalvo discusses

this logic in the context of videogames culture, although I argue that it is a representative example of reactionary logics that function more broadly:

> The rage we see expressed by threatened individuals and groups seems to be based on at least two factors—sexist (as well as racist, homophobic and ageist) beliefs about the abilities and proper place of female players, and fears about the changing nature of the game industry. (…)
>
> some players are explicit in their complaints that growth in some areas—such as casual and social games, which are often targeted to women—means that fewer budgets and development teams will be focused on traditional titles and genres such as First Person Shooters and Action games. One component underlying this concern relates to the platforms on which such games run—meaning that hardware development and how companies like Sony and Nintendo choose to design their consoles have important implications for the games that can or cannot be developed for them. Microsoft and Sony continue to promote the graphical and computational superiority of their Xbox 360 and PlayStation 3 systems respectively. Nintendo has come under attack since at least the release of the Wii for "dumbing down" what a console could be, and (by association) for shrinking demand for potential AAA game titles (AAA games are generally considered high quality games made by large studios with big budgets). Although Nintendo wished to broaden its audience to include lapsed, older and female game players, traditional console players saw the move as one actively excluding them, and reacted quite negatively to that perceived slight. "They" would have fewer games available to play, because those games would not be available (or made for) the Wii. If they did not appear on other consoles or players chose not to buy them, the games effectively would not exist. (Consalvo 2012, 3)

Thus, a market-based decision to expand into new demographics is interpreted as an attack on the interests of current fans, because resources are being diverted away from 'their' games. Worse, this attack comes as an invasion from 'outside' the groups established as having safe, default 'ownership' of the space. We can see similar conflicts playing out fractally in many different contexts and in different scales: changes ranging from homosexual representation to the broadening of economic markets in pursuit of profit are framed as attacks on what are 'their' spaces.

And how should such attacks be resisted? With free-speech—at least, free-speech understood through the expansive and permissive lens provided by technolibertarian ideals (Turner 2006). The internet and gaming culture are full of those who argue that free-speech is the most slippery of slippery slopes and that any restrictions whatsoever cannot be trusted and

must be, once again, actively resisted. Any speech act is viable and valuable, particularly online (Citron 2014, 190), since it can just be responded to with more speech. Adherents to this view frame the internet as a lawless and unregulated 'Wild West,' and any attempts to change that are seen as threatening the internet's fundamental existence (Citron 2014, 19, 26). There is no nuance here, nor awareness or sympathy for how uneven the playing field could be if someone starts getting messages from dozens or hundreds of people telling them to shut up or worse. The answer from free-speech hardliners loops back to the idea that anyone who does not like it should leave the fandom/community/forum/context it happens in, or just not voice opinions that people do not like.

It is telling that the staunchest advocates for permissively interpreted free-speech have no interest in all of the measurable ways that women and marginalised groups are silenced by the harassment associated with the status quo:

> Instead, many of these crusaders claim that abusers are simply exercising their rights and that women should stop complaining. To sum up this attitude: *Free speech is okay for me, but not for you! Shut up or I'll put your life in danger by posting your home address on Twitter.* (Golding and Van Deventer 2016, 99)

As Whitney Phillips argues, this position

> calls attention to the ugly side of free-speech, which so often is cited by people whose speech has always been the most free—namely straight white cisgendered men (i.e., men whose gender identity aligns with cultural expectations for their biological sex)—to justify hateful behaviour towards marginalized groups. In these cases, claims to protected speech are often less about the legal parameters of the First Amendment and more about not wanting to be told what to do, particularly by individuals whose perspective one doesn't respect.
>
> Just as it places assumptions about free speech in a new and perhaps uncomfortable light, trolling also reveals the destructive implications of freedom and liberty, which, when taken to their selfish extreme, can best be understood as "freedom for *me*," "liberty for *me*," with little to no concern about how these actions might infringe on others' freedoms. (Phillips 2015a, 133)

Danielle Keats Citron has vigorously debunked the idea that there is any legal basis for extreme free-speech protections even under the First Amendment defences enshrined in US law (Citron 2014, 190–225). Far from being a slippery slope, she argues that the internet can never reach its potential until it is free from groups using the permissiveness of the environment to silence others (Citron 2014, 196). Of course, a central part of the issue here is that the problematic parts of cultures on and offline simply do not care about facts, particularly when presented by a woman. What matters more than facts to online cultures is being on the right side of yet another zero-sum cultural battle where invaders from 'outside' are trying to force a change to 'their' culture.

The people who enjoy what Whitney Phillips calls 'a position of pure privilege' in online and gaming spaces have noticed that demographics are changing around them, and are fighting to retain their privileged position against changes that they believe can only erode it (Phillips 2015a, 26). Writing about the dynamics surrounding what she calls 'subcultural-trolls,'[5] Phillips argues that this privilege manifests across a broad range of the assumptions which underlie trolling behaviour online:

> Needless to say, the power dynamic between the troll and his or her target is, and can only ever be, fundamentally asymmetrical. Trolls don't mean, or don't have to mean, the abusive things they say. They get to choose the extent to which their statements match their personal beliefs; they get to establish that they're just trolling (I complicate this notion of "just" trolling in later chapters). Targets of trolling, on the other hand, are expected to take trolls at their word, and are only trolled harder if they resist. Consequently, trolls exercise what can only be described as pure privilege—they refuse to treat others as they insist on being treated. Instead, they do what they want, when they want, to whomever they want, with almost perfect impunity. To call trolling behaviors ethically and ideologically fraught would be an understatement, and is a point that must be taken into consideration—in fact, must be taken as a given—in all subsequent discussions of trolling. (Phillips 2015a, 26)

One of the reasons that this position of privilege is defended so violently is because many of the people involved are both aware of their privilege and insecure despite of it (Rubin et al. 2020).

Thus, we can see that the technolibertarian foundation of internet culture has produced recurring, overlapping themes that add together into a profound resistance to change. Firstly, since 'everyone is equal' (and

equally 'disembodied') in online and gaming spaces, intersectional dimensions to identity such as gender, race, and sexuality (among others) are argued to not apply. As such, the only people who would bring them up have an unwanted and unnecessary political agenda, and are likely to be considered inauthentic invaders from outside of the established culture. Secondly, since culture is understood as a zero-sum conflict, any victories from those seeking change come, by definition, at the cost of those invested in the existing culture. As such, anyone seeking change must be treated as a potentially existential threat. Thirdly, free-speech is a resource with which such incursions can be fought, and since—because everyone is equal in online and gaming spaces—it is equally distributed, there should be no restraint upon how it is applied to conflicts. Anyone seeking to argue this is not true marks themselves as an outsider who deserves to be driven away. Lastly, the adherents of these logics are used to operating in a context of unquestioned privilege and will attack anyone challenging that position—even through questioning it.

However, although so far we have been discussing online and gaming cultures, part of the problem that Whitney Phillips identifies is that online culture is toxic because it reflects the toxicity of our cultures more broadly and cannot be separated from them (Phillips 2015a, 11, 118; 2015b, 2015c; Shaw 2014).

There is a wide cultural toxicity and hostility towards women and anyone who does not fit the straight, cisgender, white male 'default'[6] (Banet-Weiser 2015; Buni and Chemaly 2014; Condis 2015; DePass 2016; Lien 2013; Massanari 2015, 2018; Phillips 2015a, 42–45, 164; Rubin et al. 2020; Salter 2017; Vossen 2018). Harassment of women and anyone identifiable as not a straight, white, cisgender male exists across global cultures, NOT just those in online spaces, and it thus surfaces within popular culture as a whole.[7] We can see harassment campaigns unfolding in the cultures focused around comics,[8] music,[9] sport,[10] online culture,[11] film,[12] fiction publishing,[13] television,[14] tabletop role-playing games and board games,[15] academia and teaching,[16] fandom, in general,[17] and the wider technology industry.[18] These harassment campaigns feature recurring themes in what triggers them and how the people who comprise them justify their actions.

If we limit our engagement with toxic cultures marginalising anyone who is not the cultural 'default' to online spaces, we miss the fact that online cultures are a reflection of a much broader social and cultural problem.

Having explored some of the context and philosophical reasoning underlying cultures of abuse on and offline, the next section examines the historical context of coordinated and targeted harassment campaigns.

Contextualising Coordinated Harassment Campaigns

Despite their heightened visibility in recent years, their seemingly distinctive natures and often their own claims to the contrary, it is vitally important to understand coordinated harassment campaigns in the broader context of toxic cultures on and offline. They are not special or unique: they are the most visible tips of an iceberg where toxic cultures play out in microcosm.

A simple definition[19] of a coordinated harassment campaign is where a group of people select a target or targets and try to damage their ability to live their lives. This can be through traumatising them directly or trying to destroy their reputation, get them fired and so on. The level of coordination varies wildly. Often it can be as simple as introducing a target to a community likely to wish them harm, and then everybody who wishes to within the community attacks independently. However, the level of coordination can rise to planned campaigns that develop their own leadership structures. What can start as low-effort harassment can grow more organisation as the campaign takes on a life of its own. Some campaigns can be very short, and some can be sustained for literal decades.

They are also not a new phenomenon. Danielle Keats Citron tracks a number of case studies in *Hate Crimes in Cyberspace*, including the pseudonymous example of 'Nancy Andrews' that began while she was studying law in 2005 (Citron 2014, 39). An online community of people almost-certainly comprised of her fellow students were sexually harassing and threatening her, and the abuse escalated when she took legal action against the forum administrator hosting the harassment community in 2009. The harassment was still ongoing in 2011 and had over the years included multiple systematic attempts to destroy her reputation and get her fired from several different jobs. Kathy Sierra's harassment began in 2007 simply in response to her visible online presence as a tech blogger who was a woman. The harassment left her afraid to leave her house, it left her afraid to stay IN her house after its location was posted online, and it continues in various forms to this day (Citron 2014, 35–39; Sierra 2014). In 2007, the foundation for what would become the Kiwi Farms[20] harassment community was laid, as people on 4chan began obsessively archiving the work

of Christine Weston Chandler or 'Chris-chan' in order to mock it and terrorise her (Pless 2016). In 2008, 'Anna Mayer' became targeted for abuse because she wrote a blog in her spare time about her personal issues. That harassment community followed her for many years, tracking her across several attempts to delete her identity and start over. It dedicated itself to trying to make her unemployable, to frame her as a racist and to endanger her by faking accounts in her name asking for rough, non-consensual sex (Citron 2014, 1–3). In 2008, a harassment community targeted 'Zoe Yang' for her work as a sex-positive columnist at her university newspaper, following her for years and again trying to destroy her employability, eventually driving her offline entirely (Citron 2014, 193). In 2009, a woman was attacked and raped in a home invasion after her ex-boyfriend created accounts in her name and posted her address and fake fantasies of abuse and rape to Craigslist (Citron 2014, 5–6). The year 2010 was distinguished by the 'dickwolves incident' (Consalvo 2012; Phillips 2012; Salter and Blodgett 2012). The webcomic Penny Arcade released a comic framed around a rape joke, prompting complaints from several quarters, including a rape survivor who announced an intention to boycott the Penny Arcade Expo (PAX). The comic's creators mocked the boycott and the rape survivor who suggested it. Fans of the comic proceeded to harass and/or threaten to rape anyone publicly critical of it or its comedy across several years (Stanton 2011). The comic's creators further added fuel to the fire over this period by creating and selling branded 'Team Dickwolf' merchandise inspired by the controversy, underlining how normalised rape threats and harassment were.[21]

The year 2012 is an important marker in the lamentable history of coordinated harassment, because it represents an escalation in the level of organisation applied to delivering harm. The first incident is another case where someone delegated abuse to an online community: a woman's ex-husband posted images of her and her children next to ads with her address titled 'Rape Me and My Daughters.' This prompted more than 30 men to arrive and demand sex, some of whom tried to break in, forcing the person targeted to flee her home and her state with her children (Citron 2014, 6). In a second incident, Jennifer Hepler—then a writer for BioWare—was targeted for a vast wave of sexist abuse that included direct and specific death threats against her and her family (Consalvo 2012; Crecente 2013; Golding and Van Deventer 2016, 123–24). The attacks came after the already-hostile Reddit gaming community discovered an interview from 2006 where Hepler suggested that players should be able to skip combat

in the same way that they can skip cutscenes, leading the community to conclude she was 'the cancer that was killing BioWare' (Stanton 2012). However, the example with the most visible impact on the development of harassment communities was the series of attacks on Anita Sarkeesian in response to her *Tropes vs Women in Videogames* series.

Anita Sarkeesian launched a Kickstarter in May of 2012 to raise US$6000 in support of the development of a video series where she would explore the ways that problematic tropes tied to the representation of women in videogames. It raised its funding within 24 hours and rapidly attracted the attention of hostile gaming communities across a number of social media platforms and forums (Campbell 2019; Golding and Van Deventer 2016, 102–9). They began organising against it: Kickstarter's tool for reporting inappropriate fundraising was abused by mass-reporting that claimed the project was anything from fraud to pornography to terrorism in an attempt to get it banned or defunded. Sarkeesian's Wikipedia page was vandalised with sexist, racist and anti-Semitic material,[22] including pornography, until it was locked. Her pre-existing YouTube videos were mass-reported, often also as terrorism, in an attempt to get her channel and videos taken down. Attackers began sending Sarkeesian photoshopped images where she was being raped, and one made and distributed a videogame called *Beat Up Anita Sarkeesian* where images of her face were violently assaulted, becoming bruised and bloodied. She continued to work despite the campaign, and the Kickstarter raised hundreds of thousands of dollars more than she originally asked for. Anita Sarkeesian remains an important voice within the analysis and criticism of games culture, and as of 2020, *the harassment has not stopped*. It just changed its branding and has undergone some ebbs and flows of severity during the interim, as will become clear.

The inciting incidents for Gamergate happened in 2014.[23] There are extensive and deepening resources of academic writing that explores its history and origins (Braithwaite 2016; Chess and Shaw 2015; Golding and Van Deventer 2016; Jong 2014; Kidd and Turner 2016; Mortensen 2016; Polansky 2018; Salter 2017; Vossen 2018). I will provide a contextual summary here, particularly since Gamergate and its specific actions and dynamics provide such a large number of the examples explored across the rest of the book.

Nothing sets Gamergate apart from the other examples of harassment discussed previously in the chapter: in terms of its actions, it is not special or distinctive. The Rubicon it represents is that it was done with

ambitiousness and intent by people who were looking back on everything that had gone before, using the attacks on Anita Sarkeesian as a conscious model of what they wanted to achieve and recreate ('a_man_in_black' 2014; 'Poopsock Holmes' 2014). Most of the previous campaigns had been reactive—such as the attacks on Sarkeesian's fundraiser and Jennifer Hepler, and those on anyone critical of the dickwolves incident. Gamergate took the scale, scope and organisational structure of those campaigns, particularly the one against Sarkeesian from just a few years beforehand, and combined them with the examples of pre-meditated crimes using the internet as a weapon.

In August of 2014, game designer Zoë Quinn was targeted by their ex-boyfriend, who published a groundless, counter-factual screed purposefully designed to result in harassment to several different communities likely to provide it (Golding and Van Deventer 2016, 139; Pless 2014; Polansky 2018). Leena Van Deventer and Dan Golding have described the publication as an act of intimate partner abuse delivered via the internet (Golding and Van Deventer 2016, 133). It was designed to harness exactly the same hostility and toxicity discussed through this chapter and point it at a target: Zoë Quinn is a non-binary person making games from a marginalised perspective and was vulnerable to being attacked as an 'outsider' invading 'their' territory. Quinn had already been a target for harassment in 2013 for exactly this reason. The already-hostile communities the publication was sent to were primed to believe that Quinn was a threat and deserved 'punishment' for the false crimes they were accused of.[24] Those communities reacted as the abuser hoped they would.

The entire 'Gamergate' label was created after the fact, in an attempt to gain legitimacy in the public eye and justify its prior abuse. The campaign of harassment began on 15 August 2014 as 'the Quinnspiracy' tied to several meme hashtags and it rapidly achieved enough visibility that it was being covered in the videogames press. It did not declare itself #gamergate until 2 weeks later on 27 August, where it attempted to justify its actions as being about 'ethics in game journalism.' These claims still exist despite the fact that the targets of the campaign were rarely ever journalists (Golding and Van Deventer 2016, 151; Wofford 2014), and despite the fact that there has never been any evidence presented that the ethical breaches Gamergate claims to be responding to ever existed. In fact, Gamergate proceeded to systematically attack many venues who were writing about *actual* ethical breaches in videogames, alongside those giving them serious critical consideration and analysis:

> Journalists who work alone or in a small, informal group are facing more risks. They can be pressured to write positively about a game or service lest they end up permanently out of favour. The people selling the products often have the upper hand over the people who are supposed to be critiquing them, and they know it.
>
> Yet Gamergate isn't interested in any of this. It won't take on the PR and marketing departments of their favourite companies and ask for accountability.
>
> "Gamergate targets those who dare challenge the status quo," Brendan Keogh tells us. "The simple answer in lieu of the really complex and historical one is that the journalists they think are being 'unethical' are those that are beginning to think of videogames as cultural artefacts rather than simply commercial products." (Golding and Van Deventer 2016, 154)

During this period, Gamergate was inescapable—and it never really stopped. If you were writing critically about games or studying games for years after 2014, it was part of your life because the campaign was targeting either you or people you knew. The visibility of the attacks had profound impacts in many areas, including discouraging women from studying or making games (Cox 2014):

> The women asked me how I was dealing with it, and I answered honestly. One student said she wasn't sure she wanted to make games anymore. I said, "Me neither, but I think I'm going to keep going until I'm certain I don't want to anymore." She vowed to do the same, and said how much she appreciated my comment. "You're the only reason I'm sticking around," she said.
>
> Four weeks later, she was changing her degree. "Why make games when I can make something else for people that won't threaten my life?" she said to me.
>
> I had no reply. (Golding and Van Deventer 2016, 148)

The campaign attacked people it decided to target, often additionally through attacks on their family, friends, co-workers and workplaces. Anyone speaking up against it was likewise targeted and would be attacked through their family and friends as well (Golding and Van Deventer 2016, 144–47). The goal of Gamergate—and the goal of all the less well-known and visible, though no less damaging, campaigns that targeted so many people before, during and since—aimed to silence their targets.[25] Many of the people targeted describe having been 'terrorised' or argue that the

campaigns themselves amount to terrorism—sometimes due to the literal terror threats of bombings or mass shootings deployed to shut down discussion.[26] Jessica Megarry argues that the goal of online harassment is explicitly 'creating an atmosphere online where [the target] is made to feel continually scared, threatened and vulnerable' (Megarry 2014, 51). Ashley Lynch coined the term 'crowdsourced terrorism' to describe Gamergate, a term taken up by scholars such as Katherine Cross (Cross 2015; Lynch et al. 2015), and I consider it a good phrase to describe how these campaigns function.[27]

An entire era of academic writing about games had to unfold during a constant, lowering storm of abuse and threats. Gamergate attacked academics—preferentially targeting women cited as evidence by men in articles over the men writing those articles themselves (Golding and Van Deventer 2016, 178)—and academic conferences like the Digital Games Research Association (DiGRA) as a 'conspiracy against games.' Mostly, however, it selected targets who were marginalised people. It may have begun with Zoë Quinn, but quickly metastasized: Anita Sarkeesian saw such a resurgence of harassment and death threats that she had to cancel speaking engagements, leave her home and persuade her parents to evacuate theirs for a period, despite not being a journalist or in any way connected to the 'corruption' Gamergate claimed to be focused on. Over time, it developed a list of primary targets who are well known in discussions of the Gamergate campaign, but there were an untold number of other targets who are often overlooked. These lesser-known targets are often people belonging to different subaltern groups such as women, people of colour and anyone identifiably not-white, plus LGBTQA+ people, with particularly vicious antagonism for members of the trans community.

The people targeted by Gamergate endured graphic and credible threats of death and rape; their home and work addresses being posted online, and being hacked; their friends and family members being abused at all hours; their reputations being destroyed and/or being framed as paedophiles, terrorists or animal abusers. Once addresses were posted, attackers often made specific threats or were physically found at homes and workplaces. People were driven offline to escape. And the campaign lasted for *years,* burning people out from sheer exhaustion. Despite being declared 'dead' repeatedly in the press, attacks just went on and on. As we will see, the harassment campaign never properly stopped and still grinds away in 2020 and probably beyond. It just changed its branding, in much

the same way that the campaign against Anita Sarkeesian evolved into Gamergate.

Unfortunately, Gamergate did not just evolve: it went mainstream. Steve Bannon, Milo Yiannopoulos and the other neo-Nazis and white-supremacists who prefer the more neutral branding of 'alt-right' (ADL 2016; Daniszewski 2016; Willis 2016) saw the opportunity presented by a tide of angry, violent young white men (Beauchamp and Condis 2019; Condis 2018, 2019). Bannon pivoted the online harassment blog Breitbart away from mocking gamers and towards giving them targets (Grayson 2016; Lees 2016; Sarkeesian 2019). Bannon was then invited into President Trump's cabinet. We have seen the impact of the tactics first brought to bear against Anita Sarkeesian and then refined in Gamergate surface in politics and culture around the world.[28] Elliot Rodger murdered six people in 2014 after posting a sexist 'incel' manifesto online on 4chan and has been praised as a 'saint' by some online harassment communities—with further killers directly claiming him as inspiration for their attacks on women (BBC 2018a, 2018b; Cecco 2020; Hern 2018).[29] Kiwi Farms has driven more than one person to suicide and celebrated their deaths with a counter on their website (Fogel 2018; 'lightninggrrl' 2016; Pless 2016; 'Social Justice Viv' 2016). 8chan and Kiwi Farms have been linked to multiple mass killings that were celebrated in their communities (Ambreen 2019; Hankes 2018; Neiwert 2015). The white-supremacist attacks on a Christchurch mosque and Islamic centre in Aotearoa in 2019 are merely the most recent in a series of terrorist attacks fuelled by online hate groups taking their actions offline.[30]

Gamergate never deserved special notability, not for its tactics or for its choice of targets, since none of that was new. What it did was become visible in a way previous harassment campaigns had not, and that told a wide range of bad actors that these tactics worked, and were repeatable.

Without intervention, they are never going to go away.

Where We Go from Here

There are a number of ways that all of this was avoidable, or at least mitigatable, and they are interlinked.

In the case of Gamergate, many figures involved in the attacks have highlighted the extent to which both the videogames press, game companies and game store platforms refused to engage with the problem (Golding and Van Deventer 2016, 164–67; Sarkeesian 2019). The

Gamergate community was relentlessly focused on terrorising anyone who did not fit the straight, white, cisgender male model of 'a gamer,' and all of their talk about ethics was just justification after the fact—but they *did* believe they were defending games from invaders. As Leena Van Deventer and Dan Golding have argued,

> How differently the situation might have played out if the people in powerful industry positions had all come out in force as a united front. How differently it might have gone down had they said at the start: "If you believe that this harassment campaign is okay, don't buy our games. Get out. We don't want you in our community—you're not welcome here." (Golding and Van Deventer 2016, 166)

Instead, the platforms and the game companies ignored the problem and threw individual staff members under the bus when targeted by the mob—actually taking steps backward from how BioWare handled the attacks on Jennifer Hepler, since the company publicly supported her in 2012 (Stirling 2012). Lana Polansky has argued that in the years since Gamergate, the industry has started to use the threat of harassment campaigns as a control measure against their own staff as part of systematic anti-union efforts (Polansky 2018).

Effectively, the problem is that the industry makes money from the people doing the harassment campaigns and that harassment does not impact their bottom line. Persuading the groups that are framed as authorities important to the identities at the heart of harassment campaigns to get involved in future could in theory have a substantial impact. However, as we have seen, the movement of Gamergate-style tactics into the mainstream means that often harassment campaigns are attacking on behalf of some combination of naked anti-Semitism, white-supremacy and nationalism, racism, homophobia and/or transphobia. At that point, there is no one in authority tied to those identities who would do more than cheer for the mob's progress.

The focus of this book is on two central spines: firstly, if we can understand the community structures and internal dynamics common to online harassment campaigns, we will have insights on how to prevent or limit their impact in future. Secondly, it is vital to understand the role played by the spaces that harassment campaigns unfold within, and the resources those campaigns turn into weapons: online communication platforms themselves. As will be illustrated in more detail later, the design of online

spaces already shapes the community structures and internal dynamics of online harassment campaigns, together with how they attack their targets. I argue that it is possible to change the design of online spaces and social media to limit the extent to which they can be turned into weapons, and this would have a significant impact on online culture. One of the rallying cries from Leena Van Deventer and Dan Golding is as follows:

> Maybe it's most useful not to look at a chronology of abuse to work out what's encouraging such behaviour, but rather to look at the systems surrounding this abuse. What stays the same over the years? The targets change. The harassers change. But the systems that harbour the behaviour of the harassers haven't changed enough. (Golding and Van Deventer 2016, 101)

Golding and Van Deventer are speaking here to systems like the homogeneity of white male leadership in tech companies who do not take the problem seriously. However, although that is vitally relevant, I believe their statement also holds true for the design of the *literal systems* through which online abuse is delivered.

Currently the companies which run social media spaces like Twitter, Reddit, Facebook and others have been similarly unmotivated to engage with the problem as videogame companies were for Gamergate. In this case, the problem is more foundational: they do not simply make money from the people abusing targets using their platforms, they make money from the abuse itself.

The purpose of this book is to explain what we could do differently for a better, safer internet, and why social media companies will not do so without either massive consumer pressure or external regulation.

In Chap. 2, I explore danah boyd's concept of 'networked publics' and their critical context in detail, then connect them to studies engaging with understanding online harassment. Not only are networked publics a foundational concept for understanding online communication, but they can help illuminate how the fundamental designs of online spaces are being turned into weapons used against marginalised groups. The chapter explores these dynamics through a series of examples, such as the behaviour of 'blocking' tools in different contexts, Adrienne Massanari's exploration of how Reddit's design encourages the creation of what she calls 'toxic technocultures,' and social media platforms designed so that their core economy monetises abuse directly.

In Chap. 3, 'Exploring the Overlap Between Hatemobs and ARGs,' I argue that online harassment campaigns function as autonomous alternate reality games (ARGs) with no external leader or guidance. This chapter defines the structure and context of a 'normal' ARG before exploring a series of angles in which ARGs and online hatemobs overlap.

Chapter 4, 'Gaming the Rules' explores another way that ARGs and online harassment overlap: members of both kinds of community learn both social and technological 'rules' in order to manipulate them as part of 'playing the game.' This chapter explores these dynamics in detail because they present by far the most complex examples for consideration. These examples highlight the extent to which purely technological solutions are doomed to failure because they will simply represent new systems to 'game' or otherwise miss social dimensions to the problem.

Since the previous two chapters have identified the ways that crowdsourced terrorism and the harassment communities which create it function like ARGs, Chaps. 5 and 6 explore how we can use that knowledge to combat them and their influence.

Chapter 5, 'Problematic Tools and Platform Complicity,' examines current tools being used to fight harassment and why they can cause problems or become weaponised for harassment themselves. Additionally, it unpacks the ways that social media platforms are currently unmotivated to fix these problems because they directly profit from abuse.

Chapter 6, 'Reshaping the Landscape,' is grounded in an online ethnography of people who have survived harassment campaigns. It presents options that would limit the effectiveness of harassment communities in future by treating them as ARGs.

In May 2019, a coalition of countries and technology companies met in Paris to discuss ways that the internet and social media have been used to coordinate, inspire and organise terrorist attacks, such as the attack on a Christchurch mosque and Islamic centre in Aotearoa. The concluding chapter, 'The Christchurch Call to Action Summit and What Follows,' discusses the Summit, the ways that it shows promise and the ways that it does not go far—or specifically—enough into what will happen next. It finishes with some proposals for trying to improve the status quo.

However, before we can improve the status quo, we need to understand it. Chapter 2 begins that process by exploring danah boyd's tools for understanding how the mediation provided by social network platforms and online spaces shapes culture online.

Notes

1. This book will name as few of the people directly responsible for acts of terrorism and abuse on and offline as possible, because they do not deserve recognition for their crimes.
2. Across the book, I move between referring to harassment campaigns and harassment communities because there is functionally no distinction between the two: harassment communities form harassment campaigns. I tend to refer to harassment communities when looking at social dynamics and harassment campaigns when focusing on their actions and achievements, but there is not a discrete line—as will be discussed in later chapters.
3. As Terry Flew notes, these discourses are so pervasive that early studies of the internet and digital culture bought into them and tended to focus on whether technologies were 'good' or 'bad' (Flew 2005, 21). As Adrienne Shaw and Wendy Chun argue, they lacked any nuanced engagement with the ways new technology related to and reproduced existing structures of power (Chun 2005, 30; Shaw 2014, 274).
4. Megan Condis argues that this idea of cultural 'ownership' defaulting to straight, cisgender white men makes that demographic easily recruited to active white-supremacy: people who already vehemently believe that outsiders are coming to take their videogames, for example, are easy to persuade that 'foreign cultures' are coming to steal their country and/or culture (Beauchamp and Condis 2019; Condis 2019).
5. The 'subcultural trolls' that Phillips discusses is a group with substantial demographic and philosophical overlap with online harassers and other privileged groups online, though one which should not simply be collapsed in with them (Phillips 2015a, 2). Phillips argues that 'trolling' is a term which obscures a wide variety of different forms of behaviour under one vague umbrella and which both minimises the impact on the people targeted and suggests they are somehow at fault (Phillips 2013, 2014, 2015b; Shepherd et al. 2015, 3). As such, careless and over-broad use of 'trolling' to describe any online wrongdoing does nothing but serve the strategies of bad actors online. Instead, Phillips argues we should focus on the behaviour, rather than the identity: someone engaging in violently misogynist behaviour online is a violent misogynist, for example, no matter if they claim to be doing so to troll, or to be 'ironic' (Phillips 2015a, 97; 2015b), and this is the strategy embraced in this project.
6. See also able-bodied, neurotypical, allosexual, alloromantic and a great many more besides, since the presumed 'default' excludes multitudes.
7. The citations that follow are intended as a small representative sample of relevant issues in different facets of popular culture.

8. See Asselin (2014); Berlatsky (2018); '@CJPendragon' (2018); Edidin (2012); Elbein (2018); Leon (2018); MacDonald (2016)).
9. See Mayberry (2013).
10. See NZ Herald (2016).
11. See boyd (2016); Doyle (2014); Gardiner et al. (2016); Grayson (2020); Jane (2017); Lenhart et al. (2016); Meyer and Cukier (2006); NZ Herald (2015); Shaw (2014); Taub (2016); Valenti (2016).
12. See Bordoloi (2015); Cox (2015a, 2015b); Howard (2015); Jaworski (2015); Lynch (2016a); O'Neil (2015); Rozsa (2014).
13. See Chu (2015); Cox (2015b); Flood (2015); Waldman (2015); Wendig (2018).
14. See Gunn (2013); Lane (2016).
15. See Aidley (2018); 'De Scriptorice' (2016); Kreider (2015, 2016); Matijevic (2015); Welch (2014).
16. See Campbell (2016); Flaherty (2016); Holpuch (2013); Marcotte (2013); Massanari (2018); McFarlane (2016); Nilan et al. (2015, 2–3).
17. See Berlatsky (2013); Letamundi (2012).
18. See Banet-Weiser and Miltner (2016); McEwan (2013); Salter (2017); Wood (2016).
19. Which will be complicated and explored in more depth in later chapters.
20. Not formally named Kiwi Farms until 2014 (Pless 2016).
21. This also qualifies as one of the more visible early examples of monetising harassment directly.
22. For no reason other than the racist and anti-Semitic assumptions of the people leaving the insults, which is itself telling (Golding and Van Deventer 2016, 108).
23. A comprehensive timeline of events is available here: https://www.reddit.com/r/GamerGhazi/wiki/timeline ('A Comprehensive Timeline of Gamergate, with Sources' 2016).
24. The claims in question were even debunked by the man who started the abuse in the first place, after the absence of any facts to support it became clear, and after it was too late for his correction to matter in reducing the attacks—see note for August 16, 2014 at the following resource: ('A Comprehensive Timeline of Gamergate, with Sources' 2016).
25. See boyd (2007); Citron (2014, 27, 153–55, 161, 196); Cox (2014); Cross (2014a); Frank (2014, 2015); Geiger (2016); Golding and Van Deventer (2016, 97, 99, 107–8, 117, 144–46); Jane (2014); Megarry (2014); Rubin et al. (2020); Sarkeesian (2012, 2016); Shaw (2013); Shepherd et al. (2015); Taub (2016); Walschots (2015).
26. See boyd (2007); Citron (2014, 45–50, 104); Cooper (2014); Cross (2014c, 2014b, 2015); D'Argenio (2014); 'De Scriptorice' (2016); Frank

(2015); Jane (2017, 3, 5); Kunzler (2014); Lee (2014); Marcetic (2014); Neugebauer (2014); Olson (2014); Thériault (2015).

27. Brianna Wu prefers the phrase 'emotional terrorism' (Cohen 2015) on the grounds that it is insensitive to the victims of political terror campaigns to call harassment campaigns 'terrorism.' However, I argue it is an appropriate fit: as noted, many references support calling harassment campaigns like Gamergate terrorism (Cooper 2014; Kunzler 2014; Lee 2014; Marcetic 2014; Thériault 2015). In addition, a colleague who wishes to remain anonymous said that Gamergate reminded them of life during the Northern Ireland terror campaigns. They described an awareness of heightened, immediate and constant threat which interfered with the ability to live life normally and where everyone was aware that attracting negative attention meant it was likely to overflow onto their family and friends—as would standing up for anyone else being targeted.
28. See Cross (2016); Eichenwald (2016); Flaherty (2016); Fleishman (2016); Grove (2016); Lavin (2019); Lynch (2016b); Markus (2016); McCormick (2016); Megarry (2014); Müller and Schwarz (2018); Penny (2016); Polansky (2018); Resnick (2016); Resnick and Collins (2016); Rupar (2016); Sarkeesian (2019); Taub and Fisher (2018); Wagner (2014); Walter (2016).
29. Incels have been categorised as a terrorist group by the Canadian government and Royal Canadian Mounted Police (RCMP) (Bell 2020).
30. Kiwi Farms gleefully hosted the live-streamed video of the white-supremacist terrorist attack against Christchurch mosques in 2019 and refused to take them down once challenged by New Zealand police.

References

'A Comprehensive Timeline of Gamergate, with Sources'. 2016. Reddit/GamerGhazi. 17 April 2016. https://www.reddit.com/r/GamerGhazi/wiki/timeline.

ADL. 2016. Alt-Right: A Primer About the New White Supremacy. *Anti-Defamation League*. 2016. http://www.adl.org/combating-hate/domestic-extremism-terrorism/c/alt-right-a-primer-about-the.html#.WDOPmuYrLZs.

Aidley, Katie. 2018. The Truth about Sexual Harassment and Boardgaming. *Katie's Game Corner* (blog). 20 June 2018. https://katiesgamecorner.com/2018/06/20/the-truth-about-sexual-harrassment-and-boardgaming/.

'a_man_in_black'. 2014. GamerGate and the New Misogyny. *Medium*. 16 November 2014. https://medium.com/@a_man_in_black/gamergate-and-the-new-misogyny-284bea6a8bb3#.1fi1wcukr.

Ambreen, Sam. 2019. FYI: Kiwi Farms Linked to at Least 2 Murders and 4 Suicides. *Left at the Lights* (blog). 8 March 2019. https://web.archive.org/web/

20200312021829/https://samambreen.wordpress.com/2019/03/08/fyi-kiwi-farms-linked-to-at-least-2-murders-and-4-suicides/.

Asselin, Janelle. 2014. It Happened to Me: I Received Rape Threats After Criticizing a Comic Book. *XoJane*. 25 April 2014. http://www.xojane.com/it-happened-to-me/janelle-asselin-comic-book-rape-threats.

Banet-Weiser, Sarah. 2015. Popular Misogyny: A Zeitgeist. *Culture Digitally*. 21 January 2015. http://culturedigitally.org/2015/01/popular-misogyny-a-zeitgeist/.

Banet-Weiser, Sarah, and Kate M. Miltner. 2016. #MasculinitySoFragile: Culture, Structure, and Networked Misogyny. *Feminist Media Studies* 16 (1): 171–174. https://doi.org/10.1080/14680777.2016.1120490.

BBC. 2018a. Toronto Suspect Praised 'incel' Killer. *BBC News*. 25 April 2018, sec. US & Canada. https://www.bbc.com/news/world-us-canada-43883052.

———. 2018b. How Rampage Killer Became Misogynist 'Hero'. *BBC News*. 26 April 2018, sec. US & Canada. https://www.bbc.com/news/world-us-canada-43892189.

Beauchamp, Zack, and Megan Condis. 2019. White Supremacists Are Trying to Recruit American Teens through Video Games. *Vox*. 9 April 2019. https://www.vox.com/policy-and-politics/2019/4/9/18296864/gamer-gaming-white-supremacist-recruit.

Bell, Stewart. 2020. RCMP Adding Incels to Terrorism Awareness Guide. *Global News*. 8 June 2020. https://globalnews.ca/news/7021882/rcmp-incel-terrorism-guide/.

Berlatsky, Noah. 2013. 'Fake Geek Girls' Paranoia Is About Male Insecurity, Not Female Duplicity. *The Atlantic*. 22 January 2013. http://www.theatlantic.com/sexes/archive/2013/01/fake-geek-girls-paranoia-is-about-male-insecurity-not-female-duplicity/267402/.

———. 2018. The Comicsgate Movement Isn't Defending Free Speech. It's Suppressing It. *Washington Post*. 13 September 2018. https://www.washingtonpost.com/outlook/2018/09/13/comicsgate-movement-isnt-defending-free-speech-its-suppressing-it/.

Bordoloi, Raniz. 2015. The Racism Awakens: Backlash Against the New Star Wars Trailer. *Odyssey*. 3 November 2015. http://theodysseyonline.com/uc-berkeley/backlash-star-wars-movie-force-awakens/204254.

boyd, danah. 2007. Safe Havens for Hate Speech Are Irresponsible. *Apophenia*. 26 March 2007. http://www.zephoria.org/thoughts/archives/2007/03/26/safe_havens_for.html.

———. 2016. Culture of Harassment. *Data & Society—Points*. 22 November 2016. https://points.datasociety.net/culture-of-harassment-1d999adbfac3#.xy1xj26ty.

Braithwaite, Andrea. 2016. It's About Ethics in Games Journalism? Gamergaters and Geek Masculinity. *Social Media + Society* 2 (4). https://doi.org/10.1177/2056305116672484.

Buni, Catherine, and Soraya Chemaly. 2014. The Unsafety Net: How Social Media Turned Against Women. *The Atlantic*. 9 October 2014. https://www.theatlantic.com/technology/archive/2014/10/the-unsafety-net-how-social-media-turned-against-women/381261/.

Campbell, Colin. 2019. The Anita Sarkeesian Story. *Polygon*. 19 June 2019. https://www.polygon.com/features/2019/6/19/18679678/anita-sarkeesian-feminist-frequency-interview-history-story.

Campbell, Elaine. 2016. Twitter Trolls: Time for Academics to Fight Back? *Times Higher Education*. 8 August 2016. https://www.timeshighereducation.com/blog/twitter-trolls-time-academics-fight-back.

Cecco, Leyland. 2020. Canada Police Say Machete Killing Was 'incel' Terror Attack. *The Guardian*, 19 May 2020, sec. World news. https://www.theguardian.com/world/2020/may/19/toronto-attack-incel-terrorism-canada-police.

Chess, Shira, and Adrienne Shaw. 2015. A Conspiracy of Fishes, or, How We Learned to Stop Worrying About #GamerGate and Embrace Hegemonic Masculinity. *Journal of Broadcasting & Electronic Media* 59 (1): 208–220. https://doi.org/10.1080/08838151.2014.999917.

Chu, Arthur. 2015. Sci-Fi's Right-Wing Backlash: Never Doubt That a Small Group of Deranged Trolls Can Ruin Anything (Even the Hugo Awards). *Salon*. 7 April 2015. https://www.salon.com/2015/04/06/sci_fis_right_wing_backlash_never_doubt_that_a_small_group_of_deranged_trolls_can_ruin_anything_even_the_hugo_awards/.

Chun, Wendy Hui Kyong. 2005. *Control and Freedom : Power and Paranoia in the Age of Fiber Optics*. Cambridge, MA: MIT Press.

Citron, Danielle Keats. 2014. *Hate Crimes in Cyberspace*. Cambridge, MA: Harvard University Press.

'@CJPendragon'. 2018. Recently, I Had a Run in with ComicsGate Because I Posted an Article. This Led to Several Days of Targeted Harassment That Cost Me a Good Amount of Security and Friends. I Was Told Today of a Voicemail Related to That Received by @ComicCrusaders and with His Permission I Sharepic.Twitter.Com/WP7Tq6bRVV. Tweet. *@CJPendragon* (blog). 3 September 2018. https://twitter.com/CJPendragon/status/1036753563639316480.

Cohen, Peter. 2015. Emotional Terrorism and Censorship. *IMore*. 20 February 2015. http://www.imore.com/nsfw-emotional-terrorism-and-censorship.

Condis, Megan. 2015. No Homosexuals in Star Wars? BioWare, 'Gamer' Identity, and the Politics of Privilege in a Convergence Culture. *Convergence: The International Journal of Research into New Media Technologies* 21 (2): 198–212. https://doi.org/10.1177/1354856514527205.

———. 2018. *Gaming Masculinity: Trolls, Fake Geeks, and the Gendered Battle for Online Culture. Fandom & Culture*. University of Iowa Press.

———. 2019. From Fortnite to Alt-Right. *The New York Times*. 27 March 2019, sec. Opinion. https://www.nytimes.com/2019/03/27/opinion/gaming-new-zealand-shooter.html.

Consalvo, Mia. 2012. Confronting Toxic Gamer Culture: A Challenge for Feminist Game Studies Scholars. *Ada: A Journal of Gender, New Media, and Technology*, no. 1 (November). https://doi.org/10.7264/N33X84KH.

Cooper, Ryan. 2014. How to Stop Misogynists from Terrorizing the World of Gamers. *The Week*. 2 September 2014. https://theweek.com/articles/444093/how-stop-misogynists-from-terrorizing-world-gamers.

Cox, Carolyn. 2014. Female Game Journalists Quit Over Harassment, #GamerGate Harms Women. *The Mary Sue*. 4 September 2014. http://www.themarysue.com/gamergate-harms-women/.

———. 2015a. Incredibly Peeved Men's Rights Activists Call for Boycott of Mad Max, Are Unintentionally Hilarious. *The Mary Sue*. 12 May 2015. http://www.themarysue.com/mra-to-the-max/.

———. 2015b. Whiny Babies Complain About LGBTQ+ Characters in Star Wars, Chuck Wendig Responds Excellently. *The Mary Sue*. 8 September 2015. http://www.themarysue.com/naboo-hoo-hoo/.

Crecente, Brian. 2013. Plague of Game Dev Harassment Erodes Industry, Spurs Support Groups. *Polygon*. 15 August 2013. http://www.polygon.com/2013/8/15/4622252/plague-of-game-dev-harassment-erodes-industry-spurs-support-groups.

Cross, Katherine. 2014a. What 'GamerGate' Reveals About the Silencing of Women. *Rewire*. 9 September 2014. https://rewire.news/article/2014/09/09/gamergate-reveals-silencing-women/.

———. 2014b. We Will Force Gaming to Be Free. *First Person Scholar*. 8 October 2014. http://www.firstpersonscholar.com/we-will-force-gaming-to-be-free/.

———. 2014c. Empire of Dirt: How GamerGate's Misogynistic Policing of 'Gamer Identity' Degrades the Whole Gaming Community. *Feministing*. 23 October 2014. http://feministing.com/2014/10/23/empire-of-dirt-how-gamergates-misogynistic-policing-of-gamer-identity-degrades-the-whole-gaming-community/.

———. 2015. Crowdsourced Terrorism Spurred SXSW to Cancel My Event. *TIME.Com*. 3 November 2015. http://time.com/4097144/sxsw-crowdsourced-terrorism/.

———. 2016. Are Progressives Being Played By WikiLeaks and Julian Assange? *The Establishment*. 25 October 2016. http://www.theestablishment.co/2016/10/25/are-progressives-being-played-by-wikileaks-and-julian-assange/.

Daniszewski, John. 2016. Writing about the 'Alt-Right'. *Associated Press*. 28 November 2016. https://blog.ap.org/behind-the-news/writing-about-the-alt-right.

D'Argenio, Angelo M. 2014. The Real Problem With #GamerGate Is Fear. *Cheat Code Central*. 5 September 2014. http://dispatches.cheatcc.com/1152.

'De Scriptorice'. 2016. Tabletop Gaming Has a White Male Terrorism Problem. *Latining*. 23 March 2016. http://latining.tumblr.com/post/141567276944/tabletop-gaming-has-a-white-male-terrorism-problem.

DePass, Tanya C. 2016. Harassment, Threats, and Trolling Online, Why Diversity in All Aspects of Gaming Is Vital. *Social Computing Symposium*. Microsoft Research. https://www.microsoft.com/en-us/research/video/social-computing-symposium-2016-harassment-threats-and-trolling-online-why-diversity-in-all-aspects-of-gaming-is-vital/.

Doyle, Sady. 2014. The Vicious Attacks of GamerGate Are the Norm for Women on the Internet. *In These Times*. 29 October 2014. http://inthesetimes.com/article/17296/the_vicious_attacks_of_gamergate_are_the_norm_for_women_on_the_internet.

Edidin, Jay Rachel. 2012. Geek Masculinity and the Myth of the Fake Geek Girl. *Comics Alliance*. 15 November 2012. http://comicsalliance.com/geek-masculinity-and-the-myth-of-the-fake-geek-girl/.

Eichenwald, Kurt. 2016. How Donald Trump Supporters Attack Journalists. *Newsweek*. 7 October 2016. http://www.newsweek.com/epileptogenic-pepe-video-507417.

Elbein, Asher. 2018. #Comicsgate: How an Anti-Diversity Harassment Campaign in Comics Got Ugly—And Profitable. 2 April 2018, sec. Entertainment. https://www.thedailybeast.com/comicsgate-how-an-anti-diversity-harassment-campaign-in-comics-got-uglyand-profitable.

Elwell, J. Sage. 2014. The Transmediated Self: Life Between the Digital and the Analog. *Convergence: The International Journal of Research into New Media Technologies* 20 (2): 233–249. https://doi.org/10.1177/1354856513501423.

Flaherty, Colleen. 2016. New Website Seeks to Register Professors Accused of Liberal Bias and 'Anti-American Value'. *Inside Higher Ed*. 22 November 2016. https://www.insidehighered.com/news/2016/11/22/new-website-seeks-register-professors-accused-liberal-bias-and-anti-american-values#.WDRhXeqdVgQ.twitter.

Fleishman, Cooper. 2016. #TheList: Alt-Right Donald Trump Trolls Have Found a New Way to Attack Journalists. *Mic*. 24 October 2016. https://mic.com/articles/157543/the-list-alt-right-donald-trump-trolls-harass-jewish-journalists-8chan-raid#.ksu4NJNkL.

Flew, Terry. 2005. *New Media: An Introduction*. 2nd ed. Victoria, Australia: Oxford University Press.

Flood, Alison. 2015. Star Wars Novelist Strikes Back at Gay Character Slurs. *The Guardian*. 11 September 2015. http://www.theguardian.com/books/2015/sep/11/star-wars-aftermath-novelist-chuck-wendig-strikes-back-at-gay-character-slurs.

Fogel, Stefanie. 2018. Video Game Developer Dies After Setting Herself on Fire. *Variety* (blog). 26 June 2018. https://variety.com/2018/gaming/news/chloe-sagal-death-1202858068/.

Frank, Jenn. 2014. On Leaving. *Infinite Lives*. 11 September 2014. http://infinitelives.net/2014/09/11/on-leaving/.

———. 2015. How to Attack a Woman Who Works in Video Gaming. *The Guardian*. 1 September 2015. http://www.theguardian.com/technology/2014/sep/01/how-to-attack-a-woman-who-works-in-video-games.

Gardiner, Becky, Mahana Mansfield, Ian Anderson, Josh Holder, Daan Louter, and Monica Ulmanu. 2016. The Dark Side of Guardian Comments. *The Guardian*. 12 April 2016. https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments.

Geiger, R. Stuart. 2016. Bot-Based Collective Blocklists in Twitter: The Counterpublic Moderation of Harassment in a Networked Public Space. *Information, Communication & Society* 19 (6): 787–803. https://doi.org/10.1080/1369118X.2016.1153700.

Golding, Dan, and Leena Van Deventer. 2016. *Game Changers: From Minecraft to Misogyny, the Fight for the Future of Videogames*. South Melbourne, VIC: Affirm Press.

Grayson, Nathan. 2016. From Gold Farming to Gamergate, The Gaming Ties of Donald Trump's White House. *Kotaku*. 29 November 2016. https://kotaku.com/from-gold-farming-to-gamergate-the-gaming-ties-of-dona-1789494823.

———. 2020. Twitch's Safety Advisory Council Rollout Has Been a Disaster. *Kotaku Australia*. 20 May 2020. https://www.kotaku.com.au/2020/05/twitchs-safety-advisory-council-rollout-has-been-a-disaster/.

Grove, Lloyd. 2016. How Breitbart Unleashes Hate Mobs to Threaten, Dox, and Troll Trump Critics. *The Daily Beast*. 1 March 2016. http://www.thedailybeast.com/articles/2016/03/01/how-breitbart-unleashes-hate-mobs-to-threaten-dox-and-troll-trump-critics.html?via=mobile&source=twitter.

Gunn, Anna. 2013. I Have a Character Issue. *The New York Times*. 23 August 2013. http://www.nytimes.com/2013/08/24/opinion/i-have-a-character-issue.html?_r=0.

Hankes, Keegan. 2018. Evidence of New Mexico School Shooter's Involvement in the Racist 'Alt-Right' Is Overwhelming. *Southern Poverty Law Center*. 8 February 2018. https://www.splcenter.org/hatewatch/2018/02/08/evidence-new-mexico-school-shooter%E2%80%99s-involvement-racist-alt-right-overwhelming.

Hern, Alex. 2018. Who Are the 'incels' and How Do They Relate to Toronto van Attack? *The Guardian*. 25 April 2018, sec. Technology. https://www.theguardian.com/technology/2018/apr/25/what-is-incel-movement-toronto-van-attack-suspect.

Holpuch, Amanda. 2013. Popular Science Blog Is Run by a Woman—To the Surprise of Some on Facebook. *The Guardian*. 20 March 2013, sec. Science. https://www.theguardian.com/science/us-news-blog/2013/mar/20/i-love-science-woman-facbook.

Howard, Adam. 2015. New 'Star Wars: The Force Awakens' Trailer Sparks Racial Backlash. *MSNBC*. 20 October 2015. http://www.msnbc.com/msnbc/new-star-wars-the-force-awakens-trailer-sparks-racial-backlash-0.

Jane, Emma A. 2014. Your a Ugly, Whorish, Slut. *Feminist Media Studies* 14 (4): 531–546. https://doi.org/10.1080/14680777.2012.741073.

———. 2017. *Misogyny Online: A Short (and Brutish) History*. SAGE Swifts. London; Thousand Oaks; New Delhi; Singapore: Sage Publications Ltd.

Jaworski, Michelle. 2015. 'Star Wars: The Force Awakens' Prompts Criticism of Rey for Being a Mary Sue. *The Daily Dot*. 21 December 2015. http://www.dailydot.com/geek/star-wars-force-awakens-rey-mary-sue-backlash/.

Jong, Carolyn. 2014. 'Fighting the Good Fight': GamerGate, NotYourShield, and Neo-Fascism. *Academia.Edu*. December 2014. https://www.academia.edu/10100661/_Fighting_the_Good_Fight_GamerGate_NotYourShield_and_Neo-fascism.

Kidd, Dustin, and Amanda J. Turner. 2016. The #GamerGate Files: Misogyny in the Media. In *Defining Identity and the Changing Scope of Culture in the Digital Age*, Advances in Human and Social Aspects of Technology, 117–139. IGI Global.

Kreider, Anna. 2015. This Post Is Insufferably Long, and I'm Sorry for That. *Go Make Me a Sandwich*. 10 March 2015. https://gomakemeasandwich.wordpress.com/2015/03/10/this-post-is-insufferably-long-and-im-sorry-for-that-longtw/.

———. 2016. You Say Hello. *Go Make Me A Sandwich*. 19 October 2016. https://gomakemeasandwich.wordpress.com/2016/10/19/you-say-hello/.

Kunzler, Jeff. 2014. The Gamification of Terror: Manufacturing GamerGate's Sphere of Delusion. *Tumblr*. 22 October 2014. http://designislaw.tumblr.com/post/100651306295/the-gamification-of-terror-manufacturing.

Lane, Carly. 2016. Steven Universe Artist Deletes Twitter Account After Experiencing Harassment From Fans. *The Mary Sue*. 14 August 2016. http://www.themarysue.com/steven-universe-artist-harassed-by-fans/.

Lavin, Talia. 2019. The Fetid, Right-Wing Origins of 'Learn to Code'. *The New Republic*. 2 February 2019. https://newrepublic.com/article/153019/fetid-right-wing-origins-learn-code.

Lee, Adam. 2014. #Gamergate: The New Face of Misogynist Terrorism. *Daylight Atheism*. 20 October 2014. http://www.patheos.com/blogs/daylightatheism/2014/10/gamergate-the-new-face-of-misogynist-terrorism/.

Lees, Matt. 2016. What Gamergate Should Have Taught Us about the 'Alt-Right'. *The Guardian*. 1 December 2016. https://www.theguardian.com/technology/2016/dec/01/gamergate-alt-right-hate-trump.

Lehdonverta, Vili. 2010. Virtual Worlds Don't Exist: Questioning the Dichotomous Approach in MMO Studies. *Game Studies* 10 (1) http://gamestudies.org/1001/articles/lehdonvirta.

Lenhart, Amanda, Michele Ybarra, Kathryn Zickuhr, and Myeshia Price-Feeney. 2016. Online Harassment, Digital Abuse, and Cyberstalking in America. *Data & Society*. http://datasociety.net/output/online-harassment-digital-abuse-cyberstalking/.

Leon, Melissa. 2018. Chelsea Cain Returns: 'Yeah, I'm Dead to Marvel. Trust Me.' *The Daily Beast*, 21 September 2018, sec. entertainment. https://www.thedailybeast.com/chelsea-cain-returns-yeah-im-dead-to-marvel-trust-me.

Letamundi, Andrea. 2012. The Psychology of the Fake Geek Girl: Why We're Threatened by Falsified Fandom. *The Mary Sue*. 21 December 2012. http://www.themarysue.com/psychology-of-the-fake-geek-girl/.

Lien, Tracey. 2013. No Girls Allowed: Unravelling the Story Behind the Stereotype of Video Games Being for Boys. *Polygon*. 2 December 2013. http://www.polygon.com/features/2013/12/2/5143856/no-girls-allowed.

'lightninggrrl'. 2016. I Am Being Stalked and Harassed by Kiwi Farms and SA. *Wrong Planet*. 24 March 2016. http://wrongplanet.net/forums/viewtopic.php?t=308671.

Lindbergh, Ben. 2020. The Rise and Fall (and Rise) of the Female-Fronted First-Person Shooter. *The Ringer*. 12 May 2020. https://www.theringer.com/2020/5/12/21254593/first-person-shooter-perfect-dark-20th-anniversary-female-protagonists.

Lynch, Ashley. 2016a. 'Bustin' Makes Boys Feel Sad—Why Ghostbusters Is So Hated. *Medium* (blog). 19 May 2016. https://medium.com/@ashleylynch/bustin-makes-boys-feel-sad-why-ghostbusters-is-so-hated-49e3c78cebb0.

———. 2016b. Gamers Are Still Over (but They're Not over Trump). *Medium*. 25 October 2016. https://medium.com/@ashleylynch/gamers-are-still-over-but-theyre-not-over-trump-807dde821512?source=user_profile%2D%2D%2D%2D%2D%2D%2D%2D%2D-3-.

Lynch, Ashley, Lucas J.W. Johnson, and Su-Laine Yeo Brodsky. 2015. Gamergate, Harassment & Wikipedia Panel. Capilano University, March. https://www.youtube.com/watch?v=pW9JalRjsIg.

MacDonald, Heidi. 2016. Bestselling Author Chelsea Cain Driven off Twitter by Harassment from Comics 'Fans'. *The Beat*. 26 October 2016. http://www.comicsbeat.com/bestselling-author-chelsea-cain-driven-off-twitter-by-harassment-from-comics-fans/.

Macklin, Graham. 2019. The Christchurch Attacks: Livestream Terror in the Viral Video Age. *Combating Terrorism Center at West Point* (blog). 18 July 2019. https://ctc.usma.edu/christchurch-attacks-livestream-terror-viral-video-age/.

Marcetic, Branko. 2014. #Gamergate Is Really about Terrorism: Why Bill Maher Should Be Vilifying the Gaming Community, Too. *Salon*. 23 October 2014. https://www.salon.com/2014/10/23/gamergate_is_really_about_terrorism_why_bill_maher_should_be_vilifying_the_gaming_community_too/.

Marcotte, Amanda. 2013. Men's Rights Activists Do Not Prove Their Point. *Slate*, 19 December 2013. http://www.slate.com/blogs/xx_factor/2013/12/19/occidental_college_flooded_with_false_rape_reports_from_men_s_rights_activists.html.

Markus, Bathania Palma. 2016. BUSTED: Trump-Loving Comment Trolls Pose as Sanders and Clinton Supporters to Divide Democrats. *Raw Story*. 21 May 2016. http://www.rawstory.com/2016/05/busted-trump-loving-comment-trolls-pose-as-sanders-and-clinton-supporters-to-divide-democrats/.

Massanari, Adrienne L. 2015. #Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures. *New Media & Society*, October. https://doi.org/10.1177/1461444815608807.

———. 2018. Rethinking Research Ethics, Power, and the Risk of Visibility in the Era of the 'Alt-Right' Gaze. *Social Media + Society* 4 (2): https://doi.org/10.1177/2056305118768302.

Matijevic, Paul. 2015. That Time Zak Smith Ran A Harassment Blog. *Ettin*! 2015. https://web.archive.org/web/20200409204001/; https://www.tumbex.com/ettinjiggywithit.tumblr/post/106855388993/.

Mayberry, Lauren. 2013. Chvrches' Lauren Mayberry: 'I Will Not Accept Online Misogyny'. *The Guardian*. 30 September 2013. http://www.theguardian.com/music/musicblog/2013/sep/30/chvrches-lauren-mayberry-online-misogyny.

McCormick, Rich. 2016. Palmer Luckey Is Funding Donald Trump's Internet Trolls with His Oculus Money. *The Verge*. 23 September 2016. http://www.theverge.com/2016/9/23/13025422/palmer-luckey-oculus-founder-funding-donald-trump-trolls.

McEwan, Melissa. 2013. Adria Richards Does Belong at Tech Conferences. *Shakesville*. 22 March 2013. http://www.shakesville.com/2013/03/adria-richards-does-belong-at-tech.html.

McFarlane, Anna. 2016. Academic Stalking. *The Thesis Whisperer*. 17 August 2016. https://thesiswhisperer.com/2016/08/17/academic-stalking/.

Megarry, Jessica. 2014. Online Incivility or Sexual Harassment? Conceptualising Women's Experiences in the Digital Age. *Women's Studies International Forum* 47 (November): 46–55. https://doi.org/10.1016/j.wsif.2014.07.012.

Meyer, Robert, and Michel Cukier. 2006. Assessing the Attack Threat Due to IRC Channels. In *2014 44th Annual IEEE/IFIP International Conference on*

Dependable Systems and Networks, 0:467–72. https://doi.org/10.1109/DSN.2006.12.

Mortensen, Torill Elvira. 2016. Anger, Fear, and Games: The Long Event of #GamerGate. *Games and Culture*, April. https://doi.org/10.1177/1555412016640408.

Müller, Karsten, and Carlo Schwarz. 2018. Fanning the Flames of Hate: Social Media and Hate Crime. *SSRN*, November. https://doi.org/10.2139/ssrn.3082972.

Neiwert, David. 2015. Illinois Woman with Neo-Nazi Leanings Charged in Canadian Mass Murder Plot. *Southern Poverty Law Center*. 18 February 2015. https://www.splcenter.org/hatewatch/2015/02/18/illinois-woman-neo-nazi-leanings-charged-canadian-mass-murder-plot.

Neugebauer, Cimaron. 2014. Terror Threat against Feminist Anita Sarkeesian at USU. *Standard Examiner*. 15 October 2014. http://www.standard.net/Police/2014/10/14/Feminist-speaker-cancels-appearance-at-USU-after-terror-threat.html.

Nilan, Pam, Haley Burgess, Mitchell Hobbs, Steven Threadgold, and Wendy Alexander. 2015. Youth, Social Media, and Cyberbullying Among Australian Youth: 'Sick Friends'. *Social Media + Society* 1 (2). https://doi.org/10.1177/2056305115604848.

NZ Herald. 2015. Clementine Ford Endured Vile Comments after Man Who Abused Her Online Was Fired from His Job. *NZ Herald*, 2 December 2015, sec. Lifestyle. https://www.nzherald.co.nz/lifestyle/news/article.cfm?c_id=6&objectid=11554472.

———. 2016. Female Football Presenter Hits Back at Sexist Abuse after Criticising Penalty. *NZ Herald*. 18 February 2016. http://www.nzherald.co.nz/world/news/article.cfm?c_id=2&objectid=11591385.

Olson, Dan. 2014. #GamerGate and Base Assumptions. *Medium*. 22 October 2014. https://medium.com/@FoldableHuman/gamergate-and-base-assumptions-transcript-ab8f91074ad7#.k8ednck7h.

O'Neil, Lorena. 2015. Men's Rights Activists Boycott 'Mad Max: Fury Road'. *CNN.Com*. 15 May 2015. http://edition.cnn.com/2015/05/15/entertainment/mad-max-fury-road-boycott-mens-rights-thr-feat/index.html.

Penny, Laurie. 2016. I'm with the Banned—Welcome to the Scream Room. *Medium*. 22 July 2016. https://medium.com/welcome-to-the-scream-room/im-with-the-banned-8d1b6e0b2932#.frqotoc89.

Phillips, Whitney. 2010. View of LOLing at Tragedy: Facebook Trolls, Memorial Pages and Resistance to Grief Online. *First Monday*. 23 September 2010. https://firstmonday.org/ojs/index.php/fm/article/view/3168/3115.

———. 2012. Trolling and/or/as/alongside Harassment: What I Learned from Writing (Poorly) about Dickwolves. *Fembot Collective*. 10 April 2012. http://fembotcollective.org/blog/2012/04/10/trolling-andorasalongside-harassment-what-i-learned-from-writing-poorly-about-dickwolves/.

———. 2013. A Brief History of Trolls. *The Daily Dot*. 20 May 2013. https://www.dailydot.com/via/phillips-brief-history-of-trolls/.
———. 2014. To Fight Trolls, Focus on Actions and Context. *NY Times*. 19 August 2014. https://www.nytimes.com/roomfordebate/2014/08/19/the-war-against-online-trolls/to-fight-trolls-focus-on-actions-and-context.
———. 2015a. *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture*. Cambridge, MA: MIT Press.
———. 2015b. Let's Call 'Trolling' What It Really Is. *The Kernel*. 10 May 2015. http://kernelmag.dailydot.com/issue-sections/staff-editorials/12898/trolling-stem-tech-sexism/.
———. 2015c. We're the Reason We Can't Have Nice Things on the Internet. *Quartz*. 29 December 2015. http://qz.com/582113/were-the-reason-we-cant-have-nice-things-online/.
Pless, Margaret. 2014. Eron Gjoni, Hateful Boyfriend. *Internet Famous Angry Men*. 6 December 2014. http://idledillettante.com/2014/12/06/eron-gjoni-hateful-boyfriend/.
———. 2016. Kiwi Farms, the Web's Biggest Stalker Community. *New York Magazine*. 19 July 2016. http://nymag.com/selectall/2016/07/kiwi-farms-the-webs-biggest-community-of-stalkers.html.
Polansky, Lana. 2018. Worse than Scabs: Gamer Rage as Anti-Union Violence. *Rhizome*. 30 October 2018. http://rhizome.org/editorial/2018/oct/30/worse-than-scabs-gamer-rager-as-anti-worker-violence/.
'Poopsock Holmes'. 2014. The Bad Apples of #GamerGate. *Medium*. 18 October 2014. https://medium.com/@poopsockholmes/the-bad-apples-of-gamergate-ba39f8fd485#.vqoywlh9z.
Resnick, Gideon. 2016. Breitbart Editor Milo Yiannopoulos Takes $100,000 for Charity, Gives $0. *The Daily Beast*. 19 August 2016. http://www.thedailybeast.com/articles/2016/08/19/breitbart-editor-milo-yiannopoulos-takes-100-000-for-charity-gives-0.html.
Resnick, Gideon, and Ben Collins. 2016. Palmer Luckey: The Facebook Near-Billionaire Secretly Funding Trump's Meme Machine. *The Daily Beast*. 23 September 2016. http://www.thedailybeast.com/articles/2016/09/22/palmer-luckey-the-facebook-billionaire-secretly-funding-trump-s-meme-machine.html?via=twitter_page.
Rowe, Don. 2019. The Online Cesspits Where Hate Found a Home. *The Spinoff* (blog). 19 March 2019. https://thespinoff.co.nz/media/19-03-2019/the-online-cesspits-where-hate-found-a-home/.
Rozsa, Matthew. 2014. The Racist #BlackStormtrooper Backlash Shows the Dark Side of Geek Culture. *The Daily Dot*. 2 December 2014. http://www.dailydot.com/opinion/racist-black-stormtrooper-backlash-star-wars/.
Rubin, Jennifer D., Lindsay Blackwell, and Terri D. Conley. 2020. Fragile Masculinity: Men, Gender, and Online Harassment. In *Proceedings of the 2020*

CHI Conference on Human Factors in Computing Systems, 1–14. CHI '20. Honolulu, HI: Association for Computing Machinery. https://doi.org/10.1145/3313831.3376645.

Rupar, Aaron. 2016. Troll Armies Rig Polls to Deceive You into Believing Trump Won First Debate. *ThinkProgress*. 28 September 2016. https://thinkprogress.org/trump-online-polls-first-presidential-debate-4chan-79aee6e0fc50#.h74btsq6r.

Salter, Anastasia, and Bridget Blodgett. 2012. Hypermasculinity & Dickwolves: The Contentious Role of Women in the New Gaming Public. *Journal of Broadcasting & Electronic Media* 56 (3): 401–416. https://doi.org/10.1080/08838151.2012.705199.

Salter, Michael. 2017. From Geek Masculinity to Gamergate: The Technological Rationality of Online Abuse. *Crime Media Culture* 14 (2): 247–264.

Sarkeesian, Anita. 2012. Anita Sarkeesian at TEDxWomen 2012. *TEDxTalks*. 5 December 2012. https://www.youtube.com/watch?v=GZAxwsg9J9Q.

———. 2016. On Twitter, Conspiracy Theories, and Information Cascades. *Feminist Frequency*. 22 February 2016. https://feministfrequency.com/2016/02/22/on-twitter-conspiracy-theories-and-information-cascades/.

———. 2019. Anita Sarkeesian Looks Back at GamerGate. *Polygon*. 23 December 2019. https://www.polygon.com/2019/12/23/20976891/anita-sarkeesian-gamergate-review-feminist-frequency-game-industry.

Shaw, Adrienne. 2014. The Internet Is Full of Jerks, Because the World Is Full of Jerks: What Feminist Theory Teaches Us About the Internet. *Communication and Critical/Cultural Studies* 11 (3): 273–277. https://doi.org/10.1080/14791420.2014.926245.

Shaw, Frances. 2013. Still 'Searching for Safety Online': Collective Strategies and Discursive Resistance to Trolling and Harassment in a Feminist Network. *The Fibreculture Journal*, no. 22. http://twentytwo.fibreculturejournal.org/fcj-157-still-searching-for-safety-online-collective-strategies-and-discursive-resistance-to-trolling-and-harassment-in-a-feminist-network/.

Shepherd, Tamara, Alison Harvey, Tim Jordan, Sam Srauy, and Kate Miltner. 2015. Histories of Hating. *Social Media + Society* 1 (2). https://doi.org/10.1177/2056305115603997.

Sierra, Kathy. 2014. Trouble at the Koolaid Point. *Serious Pony*. 7 October 2014. http://seriouspony.com/trouble-at-the-koolaid-point.

'Social Justice Viv'. 2016. Kiwi Farms, the Web's Biggest Community of… *Tumblr*. 13 September 2016. http://socialjusticeviv.tumblr.com/post/150364975017/kiwi-farms-the-webs-biggest-community-of.

Stanton, Courtney. 2011. Here Is a Project: Troll! Data! Analysis! *Here Is A Thing*. 8 February 2011. https://web.archive.org/web/20110211193546/http://kirbybits.wordpress.com/2011/02/08/here-is-a-project-troll-data-analysis/.

Stanton, Rich. 2012. When Is a Game Not a Game? *Eurogamer* (blog). 10 March 2012. https://www.eurogamer.net/articles/2012-03-10-when-is-a-game-not-a-game.

Stirling, Jim. 2012. BioWare Issues Statement in Support of Jennifer Hepler. *Destructoid*. 21 February 2012. https://www.Destructoid.com/bioware-issues-statement-in-support-of-jennifer-hepler-222338.phtml.

Taub, Amanda. 2016. The Guardian Study's Hidden Lesson: Trolls Reinforce White Male Dominance in Journalism. *Vox*. 13 April 2016. http://www.vox.com/2016/4/13/11414942/guardian-study-harassment.

Taub, Amanda, and Max Fisher. 2018. Facebook Fueled Anti-Refugee Attacks in Germany, New Research Suggests. *The New York Times*. 21 August 2018, sec. World. https://www.nytimes.com/2018/08/21/world/europe/facebook-refugee-attacks-germany.html.

Thériault, Anne. 2015. Let's Call Female Online Harassment What It Really Is: Terrorism. *Vice*. 12 February 2015. http://www.vice.com/en_ca/read/lets-call-female-online-harassment-what-it-really-is-gender-terrorism-481.

Turner, Fred. 2006. *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago; London: University of Chicago Press.

Valenti, Jessica. 2016. Insults and Rape Threats. Writers Shouldn't Have to Deal with This. *The Guardian*. 14 April 2016. http://www.theguardian.com/commentisfree/2016/apr/14/insults-rape-threats-writers-online-harassment.

Vossen, Emma. 2018. On the Cultural Inaccessibility of Gaming: Invading, Creating, and Reclaiming the Cultural Clubhouse. *UWSpace*. http://hdl.handle.net/10012/13649.

Wagner, Kyle. 2014. The Future of the Culture Wars Is Here, and It's Gamergate. *Deadspin*. 14 October 2014. http://deadspin.com/the-future-of-the-culture-wars-is-here-and-its-gamerga-1646145844.

Waldman, Katy. 2015. 2015 Hugo Awards: How the Sad and Rabid Puppies Took over the Sci-Fi Nominations. *Slate.Com*. 8 April 2015. http://www.slate.com/blogs/browbeat/2015/04/08/_2015_hugo_awards_how_the_sad_and_rabid_puppies_took_over_the_sci_fi_nominations.html.

Walschots, Natalie. 2015. Gamergate: The Greatest Trick The Devil Ever Pulled. *The Establishment*. 18 November 2015. https://medium.com/the-establishment/gamergate-the-greatest-trick-the-devil-ever-pulled-876aa73e3d2e.

Walter, Damien. 2016. How the Alt-Right Invaded Geek Culture. *The Independent*. 29 August 2016. http://www.independent.co.uk/voices/how-the-alt-right-invaded-geek-culture-a7214906.html.

Welch, Mary Agnes. 2014. Taking a Stand against Abuse. *Winnipeg Free Press*. 29 November 2014. http://www.winnipegfreepress.com/local/taking-a-stand-against-abuse-284204591.html?cx_navSource=d-tiles-1.

Wendig, Chuck. 2018. In Which I Am Fired From Marvel. *Terribleminds* (blog). 12 October 2018. http://terribleminds.com/ramble/2018/10/12/in-which-i-am-fired-from-marvel/.

Willis, Oliver. 2016. What Is The 'Alt-Right'? A Guide to the White Nationalist Movement Now Leading Conservative Media. *Media Matters*. 25 August 2016. http://mediamatters.org/blog/2016/08/25/what-alt-right-guide-white-nationalist-movement-now-leading-conservative-media/212643.

Wofford, Taylor. 2014. Is GamerGate About Media Ethics or Harassing Women? Harassment, the Data Shows. *Newsweek*. 25 October 2014. http://www.newsweek.com/gamergate-about-media-ethics-or-harassing-women-harassment-data-show-279736?piano_t=1.

Wood, Holly. 2016. What Do Jason Calacanis, Marc Andreessen and Vivek Wadhwa All Have in Common? *Medium*. 9 January 2016. https://medium.com/@girlziplocked/what-do-jason-calacanis-marc-andreessen-and-vivek-wadhwa-all-have-in-common-f0286f5fdbd4#.a36mgvawh.

CHAPTER 2

Networked Publics of Abuse

The design, conceptualisation and implementation of online spaces are not neutral to the cultures that form and grow within them. danah boyd introduces the concept of a 'networked public' to explore the ways that mediation impacts culture:

> Networked publics are publics that are restructured by networked technologies. As such, they are simultaneously (1) the space constructed through networked technologies and (2) the imagined collective that emerges as a result of the intersection of people, technology, and practice. (…)
>
> While networked publics share much in common with other types of publics, the ways in which technology structures them introduces distinct affordances that shape how people engage with these environments. (…)
>
> As a result, new dynamics emerge that shape participation. (…)
>
> Networked publics' affordances do not dictate participants' behavior, but they do configure the environment in a way that shapes participants' engagement. (boyd 2011, 39)

The analysis across this book is grounded in the premise that since online spaces are publics shaped by networked technologies, it is possible to assess how the affordances provided by those networked technologies impact the kind of public they produce. In this specific context, it is possible to analyse how those affordances encourage the formation of hate-mobs, or how affordances could potentially be applied by them as harassment tools. It is important to flag that boyd's framework for

K. Veale, *Gaming the Dynamics of Online Harassment*,
https://doi.org/10.1007/978-3-030-60410-3_2

understanding networked publics understands them as emergent structures: although the design of online spaces absolutely dictates what can and cannot happen within them,[1] communities respond to those designs and are shaped by them, but not dictated by them.

> This is not to say that what emerges in social network sites is simply determined by the technical affordances or that the dynamics described here predict practices. Rather, participants are implicitly and explicitly contending with these affordances and dynamics as a central part of their participation. In essence, people are learning to work within the constraints and possibilities of mediated architecture, just as people have always learned to navigate structures as part of their daily lives. (boyd 2011, 55)

Although this has always been true and is intuitive for anyone who has spent time online, it is also clear that little scrutiny is applied by social media platforms and online companies to how their designs could be abused by bad actors actively looking for tools or traps. It is very clear that these companies spend a substantial amount of time and energy considering how to design networked publics for the audiences that they intend to attract, as Amanda Friz and Robert W. Gehl have illustrated in the case of Pinterest (Friz and Gehl 2016):

> As the new user signs up, she or he is isolated from the rest of the site. Other users are not mentioned during the tutorial, nor are new users shown how to see who's following them or how many followers they have. New users cannot see who else is on Pinterest or what they have pinned until after finishing the sign-up process and tutorial. Instead, the focus is placed squarely on the neophyte's own particular interests and passions. When new users are asked to select their first pins and first boards, there is no indication of what is popular or even what new users typically select. There is only a list of broad possibilities. Current users can invite friends to join Pinterest (and indeed this was the only way to join Pinterest for the first years of its existence), but that person's interests and boards are not mentioned at all during the tutorial, nor is there any description of how to find that friend among the vast number of Pinterest users once the tutorial is over. Contrast this with, say, Facebook where users can see not only what interests their friends have liked but also how many likes a particular business, product, or hobby has received. On Pinterest, these typical avenues for competition (gathering followers, starting trends, consuming popular products, developing cliques) are downplayed or outright ignored during the tutorial. In this way, the site

> shapes expectations as much as by what it hides as by what it demonstrates. Although there are many possible ways a user can utilize Pinterest and behave while on the site, the path of least resistance, the one that is scripted into the tutorial, is one that isolates users into singular minds invested in their own personal interests. (Friz and Gehl 2016, 691–92)

Friz and Gehl argue that Pinterest's de-emphasis on competition, its attention on what are socially understood as feminine hobbies in personalised spaces, and policy documents designed to be both inaccessible and a starkly different aesthetic all combine to imply an 'ideal, feminised subject' (Friz and Gehl 2016, 699). There is nothing requiring men not to use the site, or forcing women to, but the design creates an intentional 'path of least resistance' to encourage a primarily feminine community—explicitly because that ties into the site's business model (Friz and Gehl 2016, 699). One of the central tensions that this book returns to is that despite the clear extent to which platforms consider the ways that their underlying design can intentionally shape communities for profit, there is a corresponding lack of motivation to consider how their designs encourage or shape harassment.

This chapter will explore examples of how small differences in the design of online spaces have had an impact on the communities that form within them, with particular emphasis on Adrienne Massanari's analysis of Reddit. It will then examine case studies where safety was sacrificed in pursuit of profit and compare them to examples where the business model of the platform is explicitly based on harassment.

The Politics of Networked Publics

A simple example of how a change in affordance has an impact on both culture and harassment can be seen when examining how the behaviour of 'blocking' tools work in different spaces online.

- Twitter:
  - Blocked accounts are invisible to the user blocking them.
  - User's posts are invisible to blocked accounts (As will be discussed in more depth later, blocking tools frequently only affect specific user accounts—meaning they can be easily circumvented by logging out.).

  - Blocked accounts get a message stating they have been blocked if they check the user's profile.
- Reddit:
  - Blocked accounts are invisible to the user.
  - User's posts are *visible* to blocked users.
- Pillowfort.Social:
  - Blocked accounts are invisible to the user blocking them.
  - User's posts are invisible to blocked accounts.
  - Blocked accounts cannot find the user's account profile in searches and get no message confirming a block—effectively you cease to exist to each other.
  - Users have control over whether anyone can see their posts when logged out.

On Twitter,[2] bad actors framed their 'opponents' blocking them as 'admitting they lost the argument' and would take screen captures of the block message as evidence of their victory. On Reddit, blocking someone means that if they follow you between subreddits or threads abusing you or posting your personal information, *you won't be able to see what they're doing*. As can be imagined, these distinct affordances have a large impact on the networked publics they are associated with.[3]

There is a deepening collection of work that argues the decisions involved in creating online spaces are innately political. These works explore how the design of those spaces directly shape the purposes to which they are used (Bivens and Haimson 2016; Busch and Shepherd 2014; Duguay 2016; Friz and Gehl 2016; Gehl 2015a, b; Gillespie 2010, 2015; Kennedy et al. 2016; Milan 2015; Oakley 2016; van der Nagel 2013, 2017; van der Nagel and Frith 2015). In addition to this body of work, scholars have been applying these tools to the specific questions of how the affordances of online spaces shape online harassment.

Adrienne Massanari has produced a rich and extensive analysis of the ways that Reddit's design has produced a networked public that qualifies as a 'toxic technoculture,' showing insights into how and why Reddit became a core organisational hub for Gamergate (Massanari

2015). She highlights the ways that Reddit encourages an impression that the site reflects a straight, white, cisgender male 'geek' default through default subreddits that all accounts are initially subscribed to, which reflect geek interests. As a result, these are the largest subreddits on the site. Posts and their subreddits are more likely to become visible to the site overall if they receive more upvotes and activity, which is more likely to happen in larger subreddits. This produces a dynamic by which people who do not see themselves or their views reflected in the dominant cultures of the site choose not to participate, reinforcing the existing dynamics (Massanari 2015, 9–10). Given how often subreddits that are explicitly hostile to people outside the cultural default become endorsed through popularity, such as /r/fatpeoplehate, which is dedicated to shaming people deemed overweight, or /r/niggers (banned in 2013, immediately reborn as other subreddits), that endorsement further communicates to marginalised people that Reddit is not for them. Attempts by Reddit to correct the problem by changing the default set of subreddits that new accounts were subscribed to backfired because people chose to harass and mass-downvote rather than unsubscribing to things they did not want to engage with (Massanari 2015, 10). This shows the current state of the networked public at heart: tolerant of material that drives away people from outside the presumed straight, white, cisgender male default, and actively antagonistic to material from those groups. There were design decisions which would have helped at least make these dynamics less likely, but now that they are established, the networked publics produced within them have inertia and have demonstrated a resistance to change.[4]

It is also necessary to explore how different distinct networked publics can flourish conjoined onto one primary platform, shaped by the design of third-party software. For example, Casey O'Donnell and Mia Consalvo argue that the affordances of social media tools like TweetDeck, Hootsuite and others change the underlying networked public of Twitter, including in ways that played a central role in how Gamergate organised (O'Donnell and Consalvo 2015). These tools offered ways to change the default structure of the Twitter feed through channels that only existed at the level of hashtags, rather than the more formal structures provided by forums.[5] The Gamergate community used these spaces to plan and form their

communities. The same tools allowed for the easy management of many parallel accounts, helping with the rapid rotation of accounts as each one was blocked by a given target, maintaining access and pressure. Additionally, they also made access to and surveillance of those targets easier, since it is possible to create channels entirely focused on what one person posts. O'Connell and Consalvo argue that the design of social media tools like TweetDeck contributed to the fact people could contribute to Gamergate as *play*, albeit play with terrible consequences for the people targeted (O'Donnell and Consalvo 2015, 2).

Another dimension of designing networked publics with huge consequences on the communities that form within them is commercial. Robert W. Gehl considers the role that money plays in shaping the affordances of social media and online spaces, as he compares the underlying design principles of traditional, corporate social media (CSM) spaces to those provided by alternative social media (Gehl 2015b). One of Gehl's conclusions is that as soon as advertising and online economies are one of the bedrock principles that an online space is designed around, the experience of users becomes a secondary concern (Gehl 2015b, 5–6).

Where Harassment Is the Business Model

User experience being trumped by the pursuit of profit is also visible in the multiple social media platforms and online spaces being designed *for* abuse, so that affordances optimised for harassment or non-consensual surveillance are part of the intended attraction for their audience. For example, an iOS app called 'Girls Around Me' generated a live map based on the publicly accessible Facebook and Foursquare information belonging to women in a radius around the user, allowing the user to read their interests and potentially their exact physical location.

> Okay, so here's Zoe. Most of her information is visible, so I now know her full name. I can see at a glance that she's single, that she is 24, that she went to Stoneham High School and Bunker Hill Community College, that she likes to travel, that her favorite book is *Gone With The Wind* and her favorite musician is Tori Amos, and that she's a liberal. I can see the names of her family and friends. I can see her birthday.
>
> (…) So now I know everything to know about Zoe. I know where she is. I know what she looks like, both clothed and mostly disrobed. I know her full

> name, her parents' full names, her brother's full name. I know what she likes to drink. I know where she went to school. I know what she likes and dislikes. All I need to do now is go down to the Independent, ask her if she remembers me from Stoneham High, ask her how her brother Mike is doing, buy her a frosty margarita, and start waxing eloquently about that beautiful summer I spent in Roma. (Brownlee 2012)

As John Brownlee discusses, part of the problem is that 'Girls Around Me' was not—technically—doing anything *wrong*. It was making use of publicly accessible 'application programming interfaces' (APIs) from Facebook, Foursquare and Google Maps, and using them to display information from Facebook accounts which the users could theoretically opt out of sharing. The problem is how exposed people are in practice by the rapid changes of complex privacy options on these platforms, changes which are functionally designed to discourage anyone from taking more than a cursory interest in how they work. The fact that the business models of these platforms depend on allowing advertisers and other parties to gather information on the people using them is extremely relevant: what matters is whether the information can be accessed for potential sale, not what anyone might then do with it. And 'Girls Around Me' illustrates how trivially easy it is to put different sets of information together in ways that expose women to significant risks.[6]

Burnbook is another app which places people at risk in order to profit from the resulting traffic. It is targeted at teenagers and allows users to flag other people within the app in order for the broader audience of app users to make anonymous comments about them (Graber 2015; Turner 2015).[7] It is possible to create pages for anyone, whether or not they have an account on Burnbook themselves, meaning that there is no need for them to consent to their involvement or even be aware that it is happening. All comments are anonymous, even though the people being flagged for discussion are decidedly not. The app also helps to narrow the audience to specific contexts, since you tie it to a 'community' at a given school, and people at that school will see posts related to it. This raises the chances that the person being commented on will be known to the people using it, and know them—but be unable to identify them because of anonymity. Although the app is theoretically age-limited to people 18 or older (or 17 with parental permission), Diana Graber found 'communities' for schools with children far younger than the threshold—PreK-12, middle and high schools, in the American context (Graber 2015). In a surprise to no one, mostly Burnbook was used as a hub for online bullying and harassment,

and caught broader attention after anonymous threats of school shootings at specific schools were made using the app. It is hard to see how the foundational utility and affordances of Burnbook could have been understood as anything other than encouraging harassment. The fact the app is free to download suggests that its business model is about gathering targeted information tied to the communities people create within it—specifically schools. Those users can then be sold advertising based around knowing their location, even if they are fundamentally anonymous. The ability to abuse people anonymously is the hook, and people will need to get the app themselves to check if they are being targeted. In both directions, abuse is the motivation for people to begin using it.

Peeple is an app that takes Burnbook's premise and attempts to monetise it further from several different angles. The basic premise behind Peeple is that if you know someone's phone number, you can create an account for them through the app. Once an account is created, anyone using the service can rate them and leave reviews, leading the app to be described as 'Yelp for people.' The people rated do not need to consent to having an account created for them and will be spammed with text messages to tell them about the account and any updates—again, without their consent (Dawson 2015, 2016; Perez 2016). In effect, the app is an open door that can be used to expose people to abuse from the internet, over which they have no control. Peeple goes further into directly monetising harassment than Burnbook did, since people are only able to hide 'their' reviews by creating an account on the app. Creating an account automatically agrees to terms of service that grants Peeple complete rights to use content related to 'your' reviews in any way they wish in perpetuity (Perez 2016). Worse still, the app will show reviews that users have hidden to anyone with a paid subscription, directly monetising abuse (Dawson 2016; Perez 2016). Effectively, abuse or the potential of abuse drives people onto the app, which gives the app legal protection against complaints, and those individuals most invested in harassment are core to the business model.

Social Autopsy is both a more and less extreme example of an app commercially fuelled by harassment: it is more extreme because its 'hook' is itself a tool for harassment, rather than an attempt to monetise other people's harassment. The creators attempted to raise funding for a service which would allow people to submit personal information that would publicly identify 'online bullies and trolls,' essentially functioning as a commercial doxxing[8] database. The short-sighted and unreasonable

justification for the app was a dogged insistence that the only people who would be targeted were themselves 'bad' (Cluley 2016; Harper 2016; Owens 2016). It is less extreme only in that Burnbook is clearly framing its business model around harassment on purpose, whereas the creators of Social Autopsy were wilfully ignorant of the realities for how the service could easily be abused.[9]

Beyond these examples, the worst offenders are the broad spectrum of sites discussed by Danielle Keats Citron that encourage people to submit intimate photos and videos online without the consent of those involved as 'revenge porn.' The business model of the sites includes charging people to have their imagery removed from public view. Revenge porn sites functionally encourage (or, at minimum, provide spaces to facilitate) both criminal harassment and then blackmail to hide it (Citron 2014, 25, 175–77). However, corporate social media and more mainstream online spaces also have financial motivations not to stop the harassment delivered via their networks, since it drives engagement and account creation (Golding and Van Deventer 2016, 101; Kuchera 2014; Lewis 2018, 41–43; Lynch 2016; Saitta 2015; Salter 2017; Warzel 2016). Given the different sizes of audience and scales of impact involved, their complicity is all the more serious given the greater impact their services have on social discourse online, as will be discussed in later chapters.

The design of social networks and online spaces has a substantial impact on the kinds of communities which can form within them, an emergent property that danah boyd understands as 'networked publics.' Substantial resources are invested by online companies to understand how they can shape the networked publics they become home to via the affordances they provide in order to profit from them. However, one consequence of this focus on profit is that many social networks are either tolerant of abuse or seek to use it as an enticement to use their products. Studying the affordances of online spaces and the dynamics of how people engage with them is vital for understanding the breadth, impact and techniques of online harassment, along with how it can be limited in future. The next chapter will examine harassment communities in more detail, particularly the ways they function as autonomous alternate reality games in how they adapt themselves and the affordances they are surrounded by for the purposes of abuse.

Notes

1. Lawrence Lessig's dictum that 'code is law' certainly applies here (Lessig 2006).
2. And Google+, while it was active.
3. Pillowfort is discussed as a specific case study in Chap. 6.
4. It is important to note that Reddit is a collection of networked publics that can have very different dynamics and community 'feel.' As will be discussed later, this both builds space for people to make something new and puts those new areas at risk if they come to the attention to the site overall, precisely because of its antagonism to them.
5. Much of this functionality has been folded back into Twitter itself in the intervening time, meaning that there is now less distinction between the networked publics of Twitter and social media tools like TweetDeck.
6. John Brownlee notes that the 'Girls Around Me' app was rapidly banned from Foursquare and then the Apple iOS store after he wrote about it. Until then, he argued that the app functioned as an object lesson regarding how vitally important, and invisibly undermined, privacy has become. It was an example that barely anyone would argue was *not* crossing important lines on privacy and personal safety, and the fact that everything it was doing was legal was clearly worse rather than better.
7. Anonymity is not the cause of harassment, nor is preventing anonymous engagement a solution—as will be discussed in more depth in Chap. 5.
8. Also written as doxing, this is releasing private information and making it public as a way to encourage harassment (Golding and Van Deventer 2016, 97), which will be discussed in more detail in Chap. 3.
9. Candace Owens, one of the central figures behind Social Autopsy, was later identified as part of a network of 'alt-right' influencers on YouTube by Rebecca Lewis (Lewis 2018, 10, 21, 46).

References

Bivens, Rena, and Oliver L. Haimson. 2016. Baking Gender into Social Media Design: How Platforms Shape Categories for Users and Advertisers. *Social Media + Society* 2 (4). https://doi.org/10.1177/2056305116672486.

boyd, danah. 2011. Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In *A Networked Self: Identity, Community and Culture on Social Network Sites*, ed. Zizi Papacharissi, 39–58. New York; Abingdon: Routledge.

Brownlee, John. 2012. This Creepy App Isn't Just Stalking Women Without Their Knowledge, It's A Wake-Up Call About Facebook Privacy. *Cult of Mac* (blog). 30 March. https://www.cultofmac.com/157641/this-creepy-app-

isnt-just-stalking-women-without-their-knowledge-its-a-wake-up-call-about-facebook-privacy/.

Busch, Thorsten, and Tamara Shepherd. 2014. Doing Well by Doing Good? Normative Tensions Underlying Twitter's Corporate Social Responsibility Ethos. *Convergence: The International Journal of Research into New Media Technologies* 20 (3): 293–315. https://doi.org/10.1177/1354856514531533.

Citron, Danielle Keats. 2014. *Hate Crimes in Cyberspace*. Cambridge, MA: Harvard University Press.

Cluley, Graham. 2016. Social Autopsy Wants to Expose Trolls' Real Identities - But Is That Wise? 13 April. https://www.grahamcluley.com/2016/04/social-autopsy/.

Dawson, Ella. 2015. Why "Peeple" Is Dangerous to Survivors… and, Really, Anyone. *My Business Is Generally Pleasurable*. 1 October. https://ellacydawson.wordpress.com/2015/10/01/why-peeple-is-dangerous-to-survivors-and-really-anyone/.

———. 2016. Peeple Launches Today. Here's Its Plan to Profit from Harassment. *My Business Is Generally Pleasurable*. 7 March. https://ellacydawson.wordpress.com/2016/03/07/peeple-launches-today-heres-its-plan-to-profit-from-harassment/.

Duguay, Stefanie. 2016. Lesbian, Gay, Bisexual, Trans, and Queer Visibility Through Selfies: Comparing Platform Mediators Across Ruby Rose's Instagram and Vine Presence. *Social Media + Society* 2 (2). https://doi.org/10.1177/2056305116641975.

Friz, Amanda, and Robert W. Gehl. 2016. Pinning the Feminine User: Gender Scripts in Pinterest's Sign-up Interface. *Media, Culture & Society*. https://doi.org/10.1177/0163443715620925.

Gehl, Robert W. 2015a. Building a Better Twitter: A Study of the Twitter Alternatives GNU Social, Quitter, Rstat.Us, and Twister. *The Fibreculture Journal* 26. https://doi.org/10.15307/fcj.26.190.2015.

———. 2015b. The Case for Alternative Social Media. *Social Media + Society* 1 (2). https://doi.org/10.1177/2056305115604338.

Gillespie, Tarleton. 2010. The Politics of "Platforms". *'. New Media & Society* 12 (3): 347–364. https://doi.org/10.1177/1461444809342738.

———. 2015. Platforms Intervene. *Social Media + Society* 1 (1). https://doi.org/10.1177/2056305115580479.

Golding, Dan, and Leena Van Deventer. 2016. *Game Changers: From Minecraft to Misogyny, the Fight for the Future of Videogames*. South Melbourne, VIC: Affirm Press.

Graber, Diana. 2015. Burnbook: Who's to Blame When an App Shuts Down a School? *Huffington Post*. 27 March. http://www.huffingtonpost.com/diana-graber/burnbook-whos-to-blame-when-an-app-shuts-down-a-school_b_6948750.html.

Harper, Randi. 2016. An Open Letter to Social Autopsy. *Medium*. 15 April. https://medium.com/@randileeharper/an-open-letter-to-social-autopsy-ae64fccdcfe#.rjinmzluf.

Kennedy, Jenny, James Meese, and Emily van der Nagel. 2016. Regulation and Social Practice Online. *Continuum* 30 (2): 146–157. https://doi.org/10.1080/10304312.2016.1143160.

Kuchera, Ben. 2014. Twitter Can Fix Its Harassment Problem, but Why Mess with Success? *Polygon*. 30 July. http://www.polygon.com/2014/7/30/5952135/twitter-harassment-problems.

Lessig, Lawrence. 2006. *Code: Version 2.0*. Basic Books.

Lewis, Rebecca. 2018. Alternative Influence: Broadcasting the Reactionary Right on Youtube. *Data & Society*. https://datasociety.net/output/alternative-influence/.

Lynch, Ashley. 2016. Gamers Are Still Over (but They're Not over Trump). Medium. 25 October. https://medium.com/@ashleylynch/gamers-are-still-over-but-theyre-not-over-trump-807dde821512?source=user_profile%2D%2D%2D%2D%2D%2D%2D%2D-3-.

Massanari, Adrienne L. 2015. #Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures. *New Media & Society*, October. https://doi.org/10.1177/1461444815608807.

Milan, Stefania. 2015. When Algorithms Shape Collective Action: Social Media and the Dynamics of Cloud Protesting. *Social Media + Society* 1 (2). https://doi.org/10.1177/2056305115622481.

van der Nagel, Emily. 2013. Faceless Bodies: Negotiating Technological and Cultural Codes on Reddit Gonewild. *Scan - Journal of Media Arts Culture* 10 (2) http://scan.net.au/scn/journal/vol10number2/Emily-van-der-Nagel.html.

———. 2017. From Usernames to Profiles: The Development of Pseudonymity in Internet Communication. *Internet Histories* 1 (4): 312–331. https://doi.org/10.1080/24701475.2017.1389548.

van der Nagel, Emily, and Jordan Frith. 2015. Anonymity, Pseudonymity, and the Agency of Online Identity: Examining the Social Practices of r/Gonewild. *First Monday* 20: 3. https://doi.org/10.5210/fm.v20i3.5615.

O'Donnell, Casey, and Mia Consalvo. 2015. Games Are Social/Media(Ted)/Technology too …. *Social Media + Society* 1: 1. https://doi.org/10.1177/2056305115580337.

Oakley, Abigail. 2016. Disturbing Hegemonic Discourse: Nonbinary Gender and Sexual Orientation Labeling on Tumblr. *Social Media + Society* 2 (3). https://doi.org/10.1177/2056305116664217.

Owens, Candace. 2016. Wave Goodbye to Cyberbullies and Trolls: SocialAutopsy.Com. *Kickstarter*. 12 April. https://www.kickstarter.com/projects/1968200734/wave-goodbye-to-cyberbullies-and-trolls-socialauto.

Perez, Sarah. 2016. Controversial People-Rating App Peeple Goes Live, Has a Plan to Profit from Users' Negative Reviews. *TechCrunch*. 8 March. http://techcrunch.com/2016/03/08/controversial-people-rating-app-peeple-goes-live-has-a-plan-to-profit-from-users-negative-reviews/.

Saitta, Eleanor. 2015. How to End Online Harassment. Malmö, Sweden. http://videos.theconference.se/eleanor-saitta-how-to-end-online.

Salter, Michael. 2017. From Geek Masculinity to Gamergate: The Technological Rationality of Online Abuse. *Crime Media Culture* 14 (2): 247–264.

Turner, Amy-Mae. 2015. Burnbook: What Parents Need to Know about the Controversial App. Mashable Australia. 27 March. http://mashable.com/2015/03/26/burnbook-app/#WsVUsqZO6aqA.

Warzel, Charlie. 2016. "A Honeypot for Assholes": Inside Twitter's 10-Year Failure To Stop Harassment. *Buzzfeed*. 12 August. https://www.buzzfeed.com/charliewarzel/a-honeypot-for-assholes-inside-twitters-10-year-failure-to-s?utm_term=.wn5L3aD3l#.ra6d31y3K.

CHAPTER 3

Exploring the Overlap Between Hatemobs and ARGs

The last chapter discussed ways that the design of online spaces helps shape the communities that form within them. This chapter will begin by exploring alternate reality games (ARGs), since they are communities where members move between multiple networked publics on different platforms while still retaining a shared sense of identity. As a result, the design of online spaces significantly contributes to the community dynamics of ARGs, and this is something that ARG communities are themselves aware of: it is part of what they respond to in choosing spaces to work within. The reason that ARGs make for a useful starting point for the discussions in this chapter is that functionally, there is absolutely no structural difference between an ARG and a harassment community: the only distinction lies at the level of their goals. An ARG is collectively solving problems in service of advancing a broader story that they are invested in, whereas a harassment community is collectively solving problems in service of ruining the lives of the people they target.

This chapter will explore different facets of how the modes of engagement associated with ARGs make them distinctive experiences and then examine how each of those facets is relevant to understanding how online harassment campaigns function. First, however, it is worth establishing a brief baseline of how ARGs typically operate before we begin to unpack their anatomy in more depth and then explore the ways that anatomy is relevant to understanding harassment campaigns.

K. Veale, *Gaming the Dynamics of Online Harassment*,
https://doi.org/10.1007/978-3-030-60410-3_3

Understanding Alternate Reality Games

What makes ARGs unique as a storytelling form is that they do not have a discrete and singular textual structure, or even one with a boundary. ARGs are innately transmedia texts spread across multiple media platforms, each likely to be using multiple websites and sub-texts within different media environments. All of the constituent elements which contribute to a given ARG are concealed far and wide in order to provide material for disparate subcategories of audience, who then connect it together as part of the community's engagement with the game (Veale 2012, 180). As a result, ARGs are defined by processes of affective engagement rather than textual structure, because they can be constituted by any form of textual engagement common to the internet at large. For example, when you read and respond to an email at work, the experience has a particular affective tone associated with it. However, if you were sneaking time to read an email at work that connected to an ARG you were part of, that would be a distinctive experience—and the fact you were sneaking time would be part of what set it apart. However, there is no difference in terms of the *processes of textual engagement* involved: in both cases you are reading text on screen and then responding to it to the best of your ability (Veale 2012, 180–81; 2013). To an external observer, there is no difference in what you are doing, whereas for people involved with the game, the two experiences are very distinct.

Speaking generally, the experience of an ARG begins when someone encounters something interesting or unusual enough to prompt them to explore it further, leading to further connections. Dave Szulborski refers to these as 'rabbit holes' (Szulborski 2005). An example from the world of traditional ARGs was when a patched update to the videogame *Portal* (Valve 2007) meant that if players carried clock-radios into different parts of the game world, the radios would play distorted noises. The individuals who go down one of these rabbit holes then find themselves exploring alongside other people online, trying to discover what is going on. They then work to solve the puzzles they are presented with as an individual within a wider community in order to make progress through the story, each participant contributing with their own skills. The only 'edge' to the text is what the individual members of the community collectively believe might be relevant to it: as such, it is in a constant state of flux, and there are always ongoing arguments within the community as to what might be relevant. In the context of the *Portal* puzzle, one example would be the

people who processed the code of the sounds through steganography software, uncovering that they were badly distorted ASCII images that hinted at the official announcement for *Portal 2* (Meer 2010a, b, c; Veale 2012, 209–11). Many other people were trying their own approaches to decoding the same information, effectively competing to be the best detectives. Another example comes from the *I Love Bees* ARG (S. Stewart et al. 2004), where the community had to try and discover the relevance of hidden numbers discovered in website code:

> An explosion of creative experimentation with the data ensued. Some players plotted the GPS points on a United States map in the hopes of revealing a connect-the-dot message. Others projected the earthbound coordinates onto sky maps to see if they matched any known constellations. A particularly large group collected the names of the cities to which the 210 points mapped and then tried to create massive anagrams and acrostics from them. A smaller group decided to average the two numbers in each pair of coordinates and look for an underlying statistical pattern across the set … (McGonigal 2010, 251)

The ultimate solution here was that the numbers were coordinates for payphones around the world, particularly in the United States, which would ring at the dates and times contained in the code. These calls would advance the game's narrative if people picked up the call and had the right information, and were presented as being from characters from the diegetic space of the game, scripted by the designers creating material for the player community.

Players of the early ARG *The Beast* dubbed the team of live game designers[1] who provide the puzzles of the ARG's story and adapt challenges in response to what the community of players does 'puppet masters' (McGonigal 2010, 253–54; Veale 2012, 196–202, 211). It is the puppet masters who create material and hide it in online environments in different contexts, leaving clues so it can be found. Having puppet masters produces more of the dynamics unique to the ARG experience. Firstly, they mean that it is impossible to know the end of the ARG ahead of time (J. Kim et al. 2009), because the story is co-created between the puppet masters and the player community. Secondly, the existence of puppet masters means that the community knows that there are people co-creating the text alongside them, posing challenges and adapting information in response to their actions and those of the wider community. Those people

might even be posing as members of that community as a way to introduce information to it, and there is never certainty for who they could be. In the context of normal ARGs, player communities work to avoid revealing the puppet masters 'behind the curtain,' because the members of the community are invested in not disrupting their own ability to experience the game without seeing its working parts (McGonigal 2003a, 12–13; Veale 2012, 196–98).

Having established a baseline sketch of how ARGs operate, we can consider the modes of engagement that make the experience of ARGs distinctive in more detail, and interrogate how they map onto the dynamics within harassment communities.

Rabbit Holes

As discussed earlier, rabbit holes are the first point of contact between members of the public and an ARG, and are generally points of curiosity or interest that might prompt some people to spend time exploring them and the ideas they connect to further.

For example, one entry-point into an early ARG called *The Beast* (Stewart et al. 2001) was that an individual named Jeanine Salla was given a credit as 'Sentient Machine Therapist' among the credits on the poster for the film *A.I.* (Spielberg 2001). This was an otherwise minor, small detail that would simply go unnoticed by most people. However, out of those who *did* notice, some were inspired to do more research. They discovered material online purporting to be from the year 2142 which linked Janine Salla to the apparent suicide of a man named Evan Chan, with hints that he was actually murdered. In comparison, *I Love Bees* (Stewart et al. 2004) had different entry-points for different groups of people sought for the ARG. Players who were veterans of *The Beast* were sent jars of honey mixed in with plastic letters with no explanation: one of the possible anagrams that could be made out of the letters was ILOVEBEES. A near-subliminal message of ILOVEBEES.COM was flashed up during trailers at the premiere for *I, Robot* (Proyas 2004), and after an ad for *Halo 2* (Bungie Studios 2004)—which was the product the ARG was designed to eventually promote. A third group were involved when a woman asked for help on a technical forum because of unusual glitches on her amateur-beekeeping website that began threatening her when she attempted to repair them. These varied approaches introduced several potential avenues for investigation, meaning that

- veterans of a prior ARG
- science-fiction fans who might be intrigued by the context
- fans of the *Halo* setting and
- people with technical experience and skills in coding and website design

could be drawn into the early ARG community. These groups were a substrate which it was hoped could make enough progress and gain enough visible momentum that they and the content/discussion they produced might interest other individuals who had not been exposed to one of the initial rabbit holes. As a result of the scatter-shot approach whereby entry-points into the ARG are often themselves concealed, the experience of ARGs for every individual player within the wider community begins with a moment of discovery.

In the context of harassment campaigns, sometimes the rabbit hole is a specific event, such as Anita Sarkeesian's Kickstarter to fund her *Tropes vs Women in Videogames* series, or when Zoë Quinn's ex-boyfriend published harassment bait in order to get them attacked. Unfortunately, members of what becomes a harassment community are often attracted merely by the detectable presence of anyone who is not a straight, white, cisgender man online—particularly if they have any opinion whatsoever regarding technology or popular culture (Alexander 2016; Asselin 2014; Buni and Chemaly 2014; Gardiner et al. 2016; Hess 2014; Kuznekoff and Rose 2013; Massanari 2015, 5; Phillips 2015, 42, 164; Sierra 2014; Sinders 2015; Taub 2016; Vossen 2018). As a result, there is no way for women and members of marginalised groups to avoid potentially attracting attention without taking steps to conceal their identity online.

For example, Jessica Price was targeted for harassment as a result of a chain of events that started with her giving advice for character design for Massively Multiplayer Online Games over Twitter in July 2018. She highlighted the basic condescension and 'mansplaining' at work when a videogame streamer felt entitled to correct her about work she does professionally. The entire exchange became a series of rabbit holes that spread and recirculated across social media, and Reddit threads formed to raise the visibility of how 'unreasonable' Price was being. A harassment campaign targeted Price through her employer, ArenaNet. Price was fired alongside a co-worker named Peter Fries who spoke up on her behalf when ArenaNet sacrificed them in a naked and desperate attempt to retain the goodwill of the harassment community (Smith 2018). Just two months later, Riot

Games followed exactly the same pattern by firing Daniel Klein and Mathias Lehmen. Klein defended pro-diversity events from the notoriously toxic *League of Legends* community and Lehman supported him, all of which prompted another storm of harassment (Castello 2018).[2]

As with Peter Fries and Mathias Lehmen, anyone who publicly supports those under attack is creating new rabbit holes that can see them targeted themselves. Unsurprisingly, the people more likely to be targeted for harassment are women and members of marginalised groups, while white cisgender men are often ignored for making the same statements. When white cisgender men are critical of harassment communities, harassment campaigns attack women they mention briefly as a higher priority than targeting the men themselves (Golding and Van Deventer 2016, 10–11, 178–79).

Another dynamic within harassment campaigns is that they produce their own rabbit holes for distribution to a broad audience, in an attempt to bring in new members. For example, Mahli-Ann Rakkomkaew Butt and Thomas Apperley have discussed 'Vivian James,' a cartoon mascot created to represent all of the women who allegedly supported Gamergate's goals.[3] These kinds of images circulate as rallying banners and pro-movement propaganda within a harassment community, but are also outward facing in that they invite outsiders to investigate what these memes connect to, and essentially function as advertising to new potential members. Some serve an additional purpose in attacking targets of the community and have the potential to move beyond the harassment community itself into the mainstream. For example, Gamergate targeted videogame critic Veerender Jubbal in retaliation for his activism against both harassment campaigns and sexism in the gaming industry, and because he is a Sikh and thus outside the white 'default.' A recurring theme in Gamergate's harassment campaign against Jubbal involved photoshopping images of him to appear as a terrorist. Specific images that framed him for involvement in the Paris terrorist attacks of 2015 were published in the mainstream news, implying he was a suspect in the articles connected to them and putting him at risk from police forces (Cox 2015; Jubbal 2015; Lum 2015; Mastroianni 2015; Stanton 2015).

The Goal of the Game

In ARGs, the goal varies with time, but is likely to follow trends at an individual level and a community level. At an individual level, there is the initial movement to understand what is going on underneath the rabbit hole and then to overcome the obstacles between you/the community and more information. As a clearer understanding of the overall picture develops, the goal for the community is to answer the mystery or resolve the storyline—such as solving the puzzle of the mysterious noises in the *Portal* update or the way that the community of players rallied for the rights of artificial intelligences in *The Beast* (McGonigal 2003b, 6; S. Stewart et al. 2001).

We can see this overall movement in *I Love Bees* (S. Stewart et al. 2004), as discussed earlier: the experience of initial players is marked by their curiosity about whichever rabbit hole they encountered, and becoming introduced to a community of other people also trying to answer similar questions. Over time, the rabbit holes become less relevant than the immediate challenges, debates and factions within the community, which are themselves shaped by new material and developments from the puppet masters.

For online harassment campaigns, the specific goals can be quite varied but have a single unifying theme: make life, particularly life online, untenable for the people targeted (Citron 2014, 113). The Kiwi Farms harassment community has a specific goal of driving its targets to suicide (Ambreen 2019; lightninggrrl 2016; Pless 2016; Social Justice Viv 2016).

The process has a similar life-cycle to a 'normal' ARG: individuals encounter material that they consider worth following up. This could just be the visible presence of someone outside the 'default' they accept as normal or someone who does not 'deserve' being a credible voice (Sierra 2014). The individuals so inspired either respond directly themselves or look for other people already doing so that they can join. We can see this behaviour in how the harassment campaign targeting Jessica Price formed: the Youtube streamer quoted her comments about his behaviour to his followers, priming a percentage of them to take up arms on his behalf.[4] Parts of the harassment community begin 'dogpiling' onto the target, overloading their ability to use social networking and/or email with abusive messages. The disparities of scale involved mean that the size of the community accomplishes flooding the targets with vitriol even if members contribute little at an individual level (Geiger 2016).[5] The campaign seeks

out private information via asymmetrical information gathering that can then be used to find points of vulnerability, such as the target's workplaces as an avenue to getting them fired or family members (Massanari 2018, 3–4). Threats of rape and murder will be made across social media and also delivered to physical locations like homes and workplaces once they are discovered—together with threats against friends, family members and supporters once they are identified (Romano 2014; Sinders 2015). Anything that can be done to paint the target as unreliable and/or criminal will be a priority, and vast amounts of content will be created to fill the broader media landscape with the idea that the target is somehow fabricating their abuse.

A high-priority goal is to silence the targets of the harassment campaign by making them either unwilling to risk more attacks or by driving them offline (boyd 2007; Citron 2014, 27, 153–55, 161, 196; Cox 2014; Cross 2014; Frank 2014, 2015; Geiger 2016; Golding and Van Deventer 2016, 97, 99, 107–8, 117, 144–46; Jane 2014; Megarry 2014; Sarkeesian 2012, 2016; Shaw 2013; Sierra 2014; Shepherd et al. 2015; Taub 2016; Walschots 2015).

Puppet Masters

The puppet masters of traditional ARGs are the team who sets up the initial premise of the game and hides material for players to later find. They are also the people who have to respond, in real time, to the incredible problem-solving ability that a collective-intelligence of this nature can deploy. For example, the puppet masters of *The Beast* set up three months of initial puzzles before the game went live, thinking that would give them a buffer with which to create more content as the players worked. However, they found the ARG community completed that buffer within one day (J. Kim et al. 2009; McGonigal 2003b, 2–3; Meifert-Menhard 2013, 161; Sheldon 2010, 277) and had to create new content 'live' effectively in conversation with the community. A specific example of this kind of 'live' development came in *I Love Bees*. The puppet masters believed that the players would sympathise with one of the artificial-intelligence characters in the game's diegesis, dubbed 'The Sleeping Princess.' However, a group of players concluded that 'The Sleeping Princess' was a villain in disguise and betrayed her location to her enemies during a live event (J. Y. Kim et al. 2010, 41). As a result, the puppet masters had no alternative but to

include this new element into the unfolding game and rewrite what they had planned around it.

Online harassment campaigns lack puppet masters as they are traditionally understood in the context of ARGs.[6] Instead, harassment campaigns functionally place the people the community targets for abuse in the role without their consent. Their targets are the ones reacting to what the community does and introducing new challenges as they try to avoid having their lives ruined. These kinds of actions can be as simple as locking down your accounts to limit who can access them—Anita Sarkeesian deleted years worth of her online life to prevent it falling into the wrong hands (Golding and Van Deventer 2016, 109). Some people choose to document the harassment campaign to collect evidence about its activities, such as where Zoë Quinn spent time archiving the Internet Relay Chat (IRC) Chatrooms where proto-Gamergate was planning against them, before revealing those logs publicly (Futrelle 2014a, b).

Any actions that the harassment campaign's targets take in response to the harassment will be taken as somewhere between a challenge and an affront—justifying further harassment. In particular, harassment campaigns very energetically try to discredit any claims or statements made by their targets, often claiming that they are lying about their harassment or faking it themselves (Campbell 2019; Golding and Van Deventer 2016, 86–87; McWhertor 2014; Sierra 2014; Stuart 2014). Ironically, one of the fastest tests to check for a harassment campaign is to make posts talking about someone's harassment in a networked public you suspect harassment is happening within. Often harassers name-searching their targets will provide the proof themselves by attacking you and denying any harassment is happening.

Texts with No Boundaries

ARGs are distinctive partly because they never *stop*, running at all hours, and they extend technological filaments into the lives of the people involved in the community: players can invite contact using social media and email so that the game and its members can actively reach out to them as they live their normal lives (Veale 2012, 185–88). As covered earlier, the experiential dimensions of taking part in an ARG while working at your job or doing university work are very distinctive, even if the functional behaviours involved do not change. There is also an awareness that any time you are asleep or unable to check for new updates, that the

community and its discussions, challenges and debates will have moved on without you. This can produce a more specific version of the fear-of-missing-out common to social media more broadly.

Harassment communities share these dimensions too, meaning the same dynamics that see people checking social media immediately on waking up to *see what they have missed* apply—except in this case they are checking to see if there are further developments regarding harming their chosen targets.[7] And just in the same way as there are people at work, or at school, or on their phone in a room with family members whose attention is actually focused on the labour involved with an ARG, seemingly normal people are dedicating their attention to harassment communities while handling other mundane tasks.

The biggest way that harassment communities lack boundaries, however, is the way they turn these dynamics inside out and force them on the people they target without their consent. Harassment communities forcibly use technology to extend their influence into the lives of those they attack, and their only boundaries consist of what the members consider irrelevant to damaging their target. This means that the people targeted by harassment communities are aware that the ongoing mob will be running at all hours of every day and will be looking for any and all means to reach into their lives. This includes contacting/harassing their immediate family members, neighbours, workplaces or entirely unrelated people who are mistaken for the above at all hours. SWATTing—false emergency calls to encourage police to raid a target's home, and which often suggest the target is armed and a danger to other civilians, often in the middle of the night—is an example of this kind of behaviour that has received broader attention outside of harassment communities, but is a popular tactic for them as well (Cross 2015; Sinders 2015). SWATTing is home invasion by proxy, and the methods of setting them up to present the target as dangerous to police and citizenry is an attempt to delegate violence or death to the police.

The DIY Principles and Community Ethos of Harassment

One of the distinctive dimensions to engaging with ARGs is that the players must approach every challenge with their own skillset: any puzzle or obstacle has to be solved with what you personally know and can

accomplish, since there is no way to abstractly allocate resources to increasing a skill like in a videogame (Veale 2012, 182–85). If you are convinced that key information is concealed in a given email address for example, you need to either be able to hack it yourself, learn how to do so, or locate someone who is capable of doing so in the community and persuade them it is a good idea. As a result, since ARGs are too complex to be solved by one person's set of skills, they innately involve community dynamics and senses of personal accomplishment for one's contributions to the work being done.

Christy Dena introduces the term 'tiers' to describe how ARG communities stratify around different levels of engagement (Dena 2008, 42–43; 2009, 239–58). In broad strokes, the members of the primary tier are the most active members of an ARG, who bring in new material; the secondary tier fits that material together and the tertiary tier forms an audience that engages with the output of the other tiers. People move between tiers as their levels/types of engagement fluctuate. An example created by a mistake at a live event for *The Beast* showcases how tiers can work: a key prop containing part of a long password was mistakenly taken home by an actor, so the player-base programmed and distributed a client-server password cracker to fill in the gaps (McGonigal 2003a, 12–13). The people who designed the software and came up with the idea were in the primary tier. Those who drew attention to it and distributed it in the community would be in the secondary tier, potentially alongside those who installed it and helped it work. The tertiary tier would be the people following along and excited to see the problem being solved like this. People in the tertiary tier would potentially move towards the secondary tier if they were inspired to install it themselves and join in or if they began publicising its existence to other players. This approach successfully solved the problem before the puppet masters realised an issue had even occurred—and the community never realised that had not been what they were expected to do (McGonigal 2003a, 12–13).

One way to jump to the primary tier of an ARG is attending events or otherwise engaging in ARG-related activities in physical space. Often there is an opportunistic dimension to this kind of engagement in that people who live nearby have an easier time doing so, but some people are willing to travel significant distances to participate in live events. The people who visited the payphones discussed for *I Love Bees* at the set times required would qualify, and those events became hubs for groups from the ARG community—members often carried webcams so that the online

community could participate vicariously. Another example can be found in the #Cipherhunt, an ARG following the finale of *Gravity Falls* (Hirsch 2012) that was finished in under two weeks when players found a statue hidden in Oregon after solving clues hidden all over the world (Jaworski 2016). Additionally, part of the ARG ethos where there are no boundaries except what the players decide is not relevant can also play out in physical space: people in the community have talked their way into employee-only areas in hotels or followed actors to their actual homes because they thought that was what was expected of them (McGonigal 2003a, 12, 20).

Harassment communities also exhibit all of these dynamics, except that the context of tiering adapts to a situation where the goal is committing systematic, concrete harm to someone's ability to live their life. The challenge comes from overcoming any resistance provided by the people being terrorised as they try to protect themselves and those close to them. The tertiary tier functions almost exactly as it would for a normal ARG, and is made up of people who are following the activities of the harassment community and supporting them but without participating themselves. The secondary tier of harassment communities look for opportunities to promote particular achievements and individuals from the primary tier to prominence within the community, and to capitalise on progress already established by the primary tier. That progress will encapsulate a diverse set of activities because of how wildly diverse the activities of the primary tier itself is—something also true of normal ARGs.

All of the tiers of a harassment campaign are united in their focus on causing concrete harm to the people they target: there is no meaningful distinction between online and offline spaces because the goal is to take the damage into the real world (Hoverd et al. 2020).

The individuals within the primary tier do incredible amounts of labour to forward the cause of the harassment community, regardless of what mode that labour happens to be, and are most likely to be the people who can be personally identified for their contributions: part of the motivation for the labour is to achieve social capital within the community (Butt and Apperley 2016; Veale 2013). As a result, people in harassment communities are effectively competing with each other for who can do the most harm and get the most respect from their fellow harassers—a dynamic which Kathy Sierra says fuels the worst kinds of escalation:

> The attacks on you are often less about scoring points against *you* than that they're trying to out-do *one another*. They're trying to out-troll, out-hate,

> out-awful the other trolls. That's their ultimate goal. He who does the worst wins. (Sierra 2014)

Within the primary tier of harassment campaigns are a subset of individuals willing to both threaten credible physical violence and then to carry it out (Robertson 2014; Sarkeesian 2014a), such as this example from Brianna Wu:

> I got home from a movie with my husband, and someone had sent me pictures of standing right behind me in the movie theater, just to say, hey, I know where you live. (Cornish 2019)

As discussed in Chap. 1, Elliot Rodger murdered six people in 2014 after posting a sexist 'incel' manifesto on 4chan and has been praised as a 'saint' by some online harassment communities—with further killers directly claiming him as inspiration for their attacks on women (BBC 2018a, b; Cecco 2020; Hern 2018). Incels have since been categorised as a terrorist group by the Canadian government and Royal Canadian Mounted Police (RCMP) (Bell 2020). Kiwi Farms is a community that has driven more than one person to suicide and for a time ran a counter to keep score of successful victims on their site (Fogel 2018; lightninggrrl 2016; Pless 2016). Both 8chan and Kiwi Farms have been linked to multiple mass killings that were celebrated in their communities (Hankes 2018; Neiwert 2015).

It is in this light that we have to understand the white-supremacist terrorist of the Christchurch attacks referring to the livestream of his massacre as an 'effort post' on 8chan—in contrast to a low-effort 'shitpost' (Rowe 2019). His livestream was an explicit attempt to court the social capital and approval of existing white-supremacist harassment communities online.

Members of the primary tier are willing to travel to physical locations to terrorise their targets. People within Gamergate put dead animals into Zoë Quinn's mailbox (Golding and Van Deventer 2016, 143; K. Stewart 2014). Brianna Wu and her family were driven from their home by direct, specific threats after her address was posted online (Cornish 2019; Futrelle 2014c; McWhertor 2014; Stuart 2014). When a harassment campaign 'doxxes' someone by posting their personal information online (Golding and Van Deventer 2016, 97), such as their address, it gives every tier access to their physical location—and no one knows what the most active tier will do with that knowledge. Everyone involved knows that, most

acutely the people targeted, which is why it is an effective weapon of terror called *stochastic terrorism*.

Stochastic terrorism is where someone uses mass-communication such as the internet to incite people 'to carry out violent or terrorist acts that are *statistically predictable* but *individually unpredictable*' (G2Geek 2011). When someone in an extremist or harassment community brings someone to the attention of the mob through doxing them or otherwise flagging them for attention, nobody knows exactly what will happen next—but the possibility that *something* will happen rises with the size of the community. The separation between the person 'just saying' something and the person who enacts the attack creates plausible deniability (Keats 2019), but the so-called lone wolf attackers inspired to stochastic terrorism could not function outside of the broader tiers of the extremist communities that they operate within.

> White-supremacist terrorism has what amounts to a dating app online, putting like-minded individuals together both through mainstream social media platforms and more remote venues, such as 8chan, that exist to foster rage. It is online, much like Islamic terrorism, that white supremacy finds its friends, colleagues who both validate and amplify the rage. When one of them puts the violent rhetoric into action in the real world, the killer is often called a "lone wolf," but they are not alone at all. They gain strength and solace from like-minded individuals. No one would have said an individual Klansman attending a Klan meeting in the woods was a lone wolf; 8chan and other venues are similar meeting spaces in the digital wild. (Kayyem 2019)

Exploring examples of people identifiably working in the primary tier of harassment campaigns illustrates both the diversity of activity and the amount of labour that people are willing to dedicate to causing their targets harm, and gain social capital by doing so.

As discussed in Chap. 1, people within Anonymous travelled to the grave of 13-year-old Mitchell Henderson after his suicide in order to take photographs with it and turn it into a meme, as part of terrorising his surviving family members (Phillips 2015, 28–29).

Benjamin 'Bendilin' Daniel created the *Beat Up Anita Sarkeesian* videogame as part of the harassment campaign targeting her in 2012, which was then popularised and circulated by members of the secondary tier, drawing it to the attention of the third 'audience' tier (Klee 2014; Sarkeesian 2012, 2014b).

The people who make Vivian James images for rallying points, propaganda posters and porn are also generating new content for the harassment community in ways that can get them personal recognition and respect within it (Butt and Apperley 2016; Veale 2013).

James 'Grim' Desborough concluded that the fact no academic journals published articles supporting Gamergate was evidence of their corruption and bias, and tried to create his own (Desborough 2015). This is an example of attempted work at the primary tier that dead-ended: to date, 'Popular Ludology' has not published anything, despite significant initial fanfare circulated by the secondary tier of the harassment community. People are willing to invest substantial amounts of time and energy under their own name, and are taking the risk that their efforts turn out to be a complete waste of time.[8]

The people who produced and sold videogames glorifying the 2019 terrorist attacks in Christchurch which were then banned as objectionable material by the New Zealand Office of Film and Literature Classification would also qualify as working within the primary tier (O'Connor 2019; Tait 2019).

The individuals who uncover personal information through deep research and circulate it to the harassment community are in the primary tier, as are the people who directly harass targets either in person, through physical proximity or through media like phones (Campbell 2017; Dewey 2016; McKibben 2016).

The people who search social media to identify and list dissenting voices would qualify as within the primary tier, while those targeting people identified and shared on those lists would be secondary.

The people who create fake tweets[9] or impersonation accounts to foment outrage against people like Anita Sarkeesian (Sarkeesian 2014b, 2015b, 2016) are within the primary tier, while the people who actively distribute them are in the secondary tier.

The same dynamic applies to the people who created fake 75% off coupons for Nike aimed at people of colour with QR codes that read

> This is a ROBBERY, Move slowly and put all the LARGE bills in the shoe box OR everyone DIES

when scanned (McGuill 2018), in order to put random PoC at risk of a police shooting.

When Steven Polk was unable to find examples of women who supported the Gamergate cause, he created a 'sockpuppet' named Alison Prime who became a central figure within the harassment community and the #notyourshield movement. In this case the constructed persona and social media presence was detailed enough that she existed in the primary tier, independent of the person who created her (Lynch 2015; Walschots 2015).[10] Joshua Goldberg was inspired by a similar problem to Polk, in that he wanted to produce evidence of how villainous the 'enemies' facing the Gamergate harassment community were, and after not finding any evidence, he decided to create some. Goldberg was responsible for a veritable army of sockpuppets, some of whom supported Gamergate as part of the #notyourshield campaign, while others attacked people within the harassment community. The goal here was to provide evidence of how villainous the opposition to Gamergate was and justify the impression that Gamergate were the underdogs: it was a false-flag exercise for propaganda purposes.[11] He was later arrested by the FBI and charged with domestic terrorism after creating social media accounts for fake terrorists to 'prove' how dangerous the Muslim community was (Marcotte 2015; Schubert 2015).

One of the more unusual ways that tiering came to light within the Gamergate harassment community was the revelation that much of Milo Yiannopoulos' writing in support of Gamergate on the Breitbart harassment blog was actually produced by a collection of volunteers from 4chan, and then published under his byline (Bernstein 2016; Lynch 2016).

As well as volunteers dedicating significant time to crafting individual sockpuppet personae or armies of more disposable accounts, there are also the people creating collections of literal bot accounts to amplify particular perspectives, to make a given group seem larger than it is or simply to hound targets (@bethany_lacina 2018; D'Souza 2018; Gardner 2018; Lapowsky 2018). In those cases, people developing bots would qualify as being in the primary tier, while people following their instructions for creating new bots or using their botnets would be in the secondary tier.

Tiering also helps to explain one of the more dangerous and unpredictable phenomena associated with online harassment communities: on a seemingly random cycle, people previously targeted by one harassment campaign will be swept up in a tide of renewed abuse without warning, after having been seemingly forgotten. This cycle is tied to the visibility of harassment community activity within social media and other online spaces, often connected to victories or defeats or simple publicity. Visibility

attracts the attention of people who have fallen out of the harassment community, bringing them into the more audience-like tertiary tier and potentially sweeping up new members at the same time. As the tertiary tier swells, the number of people who are inspired to climb into the secondary or primary tiers also climbs. They then attack their own preferred targets, who often have very little to do with the event that attracted the community's renewed attention in the first place.[12] An iconic example of this process would be the way Gamergate attacked Anita Sarkeesian again in 2014. She was entirely uninvolved with the events that inspired the harassment campaign or even their fabricated justifications for it. She was attacked purely because members of the community remembered targeting her in 2012 and chose to attack her again. A more recent example is that in the wake of Jessica Price's firing after ArenaNet tried to appease a harassment campaign, women across the industry reported that their employers had received emails complaining about their social media presences as well—except they often contained %FEMALENAME instead of referring to them directly (Farokhmanesh 2018). This example draws attention to another phenomenon that harassment communities share with ARGs, where people produce and share instructional guides within the community.

Instruction Manuals and Influencers for Crowdsourced Terrorism

In the context of ARGs, members of the community devote time and energy to writing 'Guides' designed to get new arrivals to the community 'up to speed' on what is going on and ways that they can get involved (Dena 2008, 50–51; Veale 2012, 190–91). Producing guides sits on the border between the primary and secondary tiers of ARG engagement, and it is possible to become renowned within the community for producing guides. The first 'Guide' produced for an ARG[13] was written by Adrian Hon for *The Beast* (Janes 2019, 22) as an attempt to provide a resource for new players and was seen as analogous to a 'walkthrough' for a videogame text.[14] However, it has been noted that Hon's Guide is much more than a map to making progress through a particular ARG:

> In this regard, it is evident how a Guide can provide narrative coherence: by providing a cause-and-effect path through the components. It is the act of narrating that renders this player-created content more than a mere 'walk-through' or gameplay resource, it is a form of artistic production, a story in

> itself. Of particular importance too, is the fact that this narrative is experiential, an authentic sharing of a personal journey through the work. (Dena 2008, 51)

Effectively, Guides work because they do not try to encapsulate and speak for the diverse (and often contradictory) experiences of the whole community: they provide an account of an individual's experiences of engaging with that community and are intended to speak to an unfamiliar audience.

The context of harassment communities diversifies the roles that guides play, but ultimately, they remain an investment of resources to encourage members of the tertiary tier to move up to the secondary or primary tiers, and/or to bring new members into the tertiary 'audience' tier.[15] The posts Zoë Quinn's ex-partner made on Reddit and 4chan in order to raise the chances of producing harassment are examples (Pless 2014) of harassment guides: they are a personalised account of events designed to bring an unfamiliar audience up to speed on why a target deserves to have their life destroyed.

Alongside this form of guide, harassment campaigns like Gamergate have been infamous for virally spreading 'conspiracy map' imagery often done in Microsoft Paint (Chess and Shaw 2015, 2016; Golding and Van Deventer 2016, 174–79; Johnston 2014; Mortensen 2016) to show the sinister connections they claim exist between their targets. These images are designed to be shared across online platforms, communicate specific ideas quickly and can also be spammed at people targeted by the harassment community as a form of 'debate.'[16]

In addition, harassment communities also make use of literal guides that function as instructions designed to get new participants in the primary or secondary tiers out and terrorising as swiftly as possible. Some examples are as follows:

- Guides from Gamergate about how to use Twitter to 'flood' targets and appear to be arguing in good faith (GamerGateOP/Twitter Flooding Instructions 2014; /u/coffeeheadphone 2014).
- How to use DARVO tactics[17] to harass while claiming the moral high ground (Freyd 1997, 2016).
- How to attack advertisers on platforms that were critical of Gamergate (Golding and Van Deventer 2016, 146; Schnier 2014).

- Instructions on how to disrupt online research into diversity in games (Allaway 2014).
- And lists of people to attack and attempt to blacklist at every opportunity (Raymond 2016).

These kinds of guides were likely involved with the campaign to get Jessica Price fired from ArenaNet. They also explain the %FEMALENAME emails targeting visible women in the game industry after the Price campaign succeeded: people were literally copy/pasting the guide into their attack emails rather than personalising them for each person targeted.

In an example of how seemingly different harassment communities overlap, serial-harasser and neo-Nazi Andrew 'weev' Auernheimer can be found posting in response to an article on the white-supremacist site The Daily Stormer, coaching other neo-Nazis to attack Alison Rapp and frame her as a paedophile as part of Gamergate's attacks (Klepeck 2016):

> When you contact these people be very respectful, act as a concerned parent, link to the pro-pedo statements she's made but obviously don't link back to DS or she'll be able to dismiss it as a white supremacist conspiracy. (Anglin 2016; 'weev' Auernheimer 2016)

The kind of guide that Auernheimer presents here is particularly distinctive and appears within harassment communities far more than in normal ARG communities: attempts to guide and steer the community, potentially in secret.

Although normal ARGs certainly have people within the primary tier trying to persuade the broader community to follow particular paths, harassment communities take this much further. Some members attempt to manipulate particular outcomes and messaging—often in response to an awareness that the actions of the community will attract bad press. The Internet Relay Chat (IRC) chatlogs where Gamergate strategised their early attacks on Zoë Quinn are fixated on the possibility of driving them to suicide, but the more 'mature' voices in the space argue that would be a bad idea since it is 'not the right PR play' (Futrelle 2014a). The chatlog also illustrates discussions of how to most effectively energise or agitate the secondary and tertiary tiers of the harassment community, along with planning other manipulations directed both inside and outside the community (Futrelle 2014b). In contrast, the style guide produced by neo-Nazi Andrew Anglin for the white-supremacist site The Daily Stormer focuses on

coaching the harassment community to help them phrase their racist attacks in ways less unpalatable to the mainstream (Feinberg 2017). In effect, these efforts collectively represent an attempt to produce a fourth 'guidance' tier as a subset within the most active, primary tier of the harassment community.

Alongside the obvious motivations to become part of a guiding tier to a harassment community, such as influencing its direction and methods, is the fact that it can be extremely profitable. Whitney Phillips refers to the people who gain income from harassment as 'chaos entrepreneurs' (Cornish 2019). The fact that achieving 'hero' status in a harassment community can be parlayed into fundraising motivates people both to move into the primary tier and to compete with each other to be more notorious harassers once there. Jay Allen and 'idlediletante' have explored the ways that central highly visible members of the Gamergate harassment community profited from their involvement (Allen 2015; idlediletante 2015). Many of the people they discuss are central to the 'alternative influencer' network that Rebecca Lewis describes profiting from abuse on YouTube in 2018[18] (D'Anastasio 2018b; Lewis 2018). The QAnon and Pizzagate conspiracies display strong tiering, and the QAnon conspiracy has been intensely profitable for the three people who started it—even as it has motivated multiple stand-offs with law enforcement and inspired an arsonist to set a devastating California wildfire (Chang 2018; Murdock 2018; Robb 2017; Zadrozny and Collins 2018).

Captivating Experiences

People can find ARGs to be deeply engrossing, potentially all-consuming experiences that, Jane McGonigal argues, can 'profoundly affect their sense of identity and purpose' (McGonigal 2003b, 1):

> "I'm going to catch myself still looking for patterns and riddles in my daily life months from now," one player posted at the conclusion of a game, describing a mindset that could easily be interpreted as paranoia. Another immersive fan wrote, "We normal, intelligent people have been devoting *outrageous* percentages of our days, weeks, months to a *game*" and described the experience of playing an immersive game as kind of loss of realworld consciousness: "You find yourself at the end of the game, waking up as if from a long sleep. Your marriage or relationship may be in tatters. Your job may be on the brink of the void, or gone completely. You may have lost a scholarship, or lost or gained too many pounds". The same player subse-

> quently published a "recovery guide" for her fellow deeply immersed players, but it is important to note that she ultimately was more interested in extending, rather than recovering from, the game play: "Now here we are, every one of us excited at blurring the lines between story and reality. The game promises to become not just entertainment, but our lives." (McGonigal 2003a, 3)

One of the dimensions of ARGs which can often be lost in the details of their specific community dynamics is that they are *fun*. One of the reasons people are willing to invest incredible amounts of time and energy to them is that the people involved find doing so satisfying and profoundly enjoyable.

Appallingly, disquietingly and unsurprisingly, all of these dynamics are equally true of online harassment campaigns.

People enjoy the experience of community and kinship in harassment campaigns and can form long-term friendships united by trying to drive someone to kill themselves. It provides people a sense of righteous purpose to strike at their 'enemies' and those of the community. Overcoming obstacles and solving problems with your own skills provides the same sense of satisfaction and pride as it does in ARGs. I have witnessed people discussing wistfully that they have dropped out of a harassment community due to school or work commitments and are missing the whole experience.

The fact that online harassment is enjoyable for members of the communities involved is uncomfortable, but also fundamental. It is part of why people who participated in one ARG or harassment campaign are so primed to return to the fold: they enjoyed being involved in their last one and are looking for excuses to 'get back into it all.'

The Disobedient Resilience of ARGs

The fact that ARGs are deeply engrossing fun means that the history of ARGs is full of occasions where the people co-creatively guiding the game alongside the players lost control of the community to at least some extent. In 2002, an ARG promoting the film *Push* (McGuigan 2009) finished in a fashion that the player-base considered unsatisfactory. The conclusion the community reached was that the ARG was not actually over, and the apparent finale was a red-herring designed to throw them off the scent of the 'real game' (McGonigal 2003b, 6). Despite the fact that the puppet

masters were no longer involved in shaping the game in response to player actions and despite the fact that no more content for the ARG existed, the players persisted—even finding unrelated material which they concluded *was* part of the ongoing ARG. Another example, also from 2002, is that ARG veterans of *The Beast* discovered a website called '8March2003.com' and concluded that the ominous use of a future date suggested involvement in an ARG. They flooded the site with visitor traffic, emailed enquiries and began researching its background—including looking into the life of whoever registered the domain. The site was edited to deny any involvement in ARGs in an attempt to correct the misunderstanding and be left alone, but this failed to gain any traction. Alongside these examples, ARGs have to be open to a lack of control on the behalf of the people running them: their co-creative nature means the puppet masters have to continually adapt around unexpected decisions made by the community, as discussed earlier in the chapter (J. Y. Kim et al. 2010, 41).

Given that traditional, benign ARGs are functionally impossible to stop from the outside if the community chooses to keep going, it is obvious that harassment communities are not going to stop themselves—something clearly visible in the history of online spaces. In fact, as already discussed, they illustrate a kind of malevolent immortality in that the people motivated to participate in one campaign are easily inspired to come back or to join a new one. Since external interventions are not an option in the current landscape of networked publics, the best alternative is to look for ways of creating spaces that harassment campaigns cannot easily operate within. However, that is a substantial challenge because they are highly adept at weaponising the design of online spaces and turning them to their own ends.

Harassment campaigns can be understood as feral, self-targeting ARGs which set their own goals, and those goals typically involve ruining the lives of the people they fixate on. This chapter illustrates that understanding them as autonomous, malevolent ARGs helps explain many dynamics of how harassment campaigns have always operated. Another way that ARGs and online harassment overlap is the way that members of the community learn both social and technological 'rules' in order to manipulate them as part of 'playing the game.' The next chapter explores these dynamics in detail because they present by far the most complex examples to consider, and highlight the extent to which purely technological solutions to online harassment are doomed to failure.

Notes

1. Those designers were an element largely missing from the experience of the *Portal* ARG, since the material was static once it was added to the game.
2. These firings happened mere weeks after it was revealed that Riot Games has a deeply toxic and sexist internal structure, which prompted representatives of Riot to claim that efforts to reform the company were under way. Firing staff standing up for diversity initiatives in that context suggests that the claims were a PR fig-leaf at best (D'Anastasio 2018a; O'Connor 2018).
3. The character featured prominently and swiftly within community-generated porn and also reflected Gamergate's gender preferences by being almost entirely silent (Butt and Apperley 2016; Cole 2018).
4. This is an example of stochastic terrorism and 'tiering,' which will be discussed later in this chapter.
5. Examples of online harassment storms are not hard to find, but as a representative example, Anita Sarkeesian collated one week of harassment over Twitter here https://feministfrequency.com/2015/01/27/one-week-of-harassment-on-twitter/ (Sarkeesian 2015a). As one might expect, content warnings for graphic descriptions of abuse and threats of violence/rape.
6. The 'puppet master' label becomes particularly ironic in this context, given the way that harassment campaigns like Gamergate justify attacking their targets based on conspiracy theories that present them as manipulating events from behind the scenes, and which suggest that a vast distributed mob of attackers are somehow the underdogs in the 'conflict.'
7. The targets and friends of people being targeted experience variations on the same theme, except they need to find out what damage has happened to themselves or the people they care for while they have been away or asleep.
8. This example is useful for illustrating that vastly more projects *attempt* work at the primary tier than necessarily achieve successful harm: the ones noted and remembered are successful, and the rest are forgotten. I was able to document this example purely through seeing the secondary tier promoting the initiative on Twitter; otherwise, it would have vanished without trace.
9. This is a significant problem that involves people taking a screen capture of a tweet, editing it using something like Photoshop, and then using the edited image as 'evidence' of what someone said in ways that can be reshared by the harassment community. The ability to edit tweets is a commonly requested feature on Twitter and would amplify the problem enormously. The pattern would be to send someone threats or other vile messages, wait for them to respond and then edit the original message to be innocuous so that it looks like the response is unjustified. Images of that

interaction would then be taken as evidence and circulated to the harassment community and so on.

10. My understanding is that members of Gamergate targeted Polk for harassment after the revelation that Alison Prime had always been a sockpuppet due to his performance of femininity online, because of the high levels of transphobia and homophobia within the community.
11. 'False flag' operations are a frequent tool of harassment campaigns, where they fake membership in groups for strategic purposes. One example is Operation NotYourShield, which invented and impersonated women and people of colour who supported Gamergate, as a literal shield against claims that Gamergate was racist and misogynist: the goal was to make it seem like Gamergate critics were shouting down or talking over members of the groups they claimed to be defending (Futrelle 2014b; Johnston 2014; Jong 2014; Lynch 2015). Another pattern is where attackers claim membership in a group they wish to target while attacking another, in an attempt to inspire other groups to retaliate. 4chan often attempts to discredit feminist groups by trying to fake activist movements around topics like 'free bleeding' or the desirability of a 'bikini bridge' (Alfonso III 2014). Another example was where 4chan faked a campaign by gay men who wanted paedophiles to gain the same social acceptance as other rainbow communities (Collins 2019; Evon 2017). People also fake attacks against people in harassment communities in order to whip up sympathy and justify aggression against the enemies of the community. Joshua Goldberg is one such example, but some people take their initiatives into physical spaces. For example, in 2018, a poster on 4chan was arrested after posting plans to attack the white-supremacist 'Unite the Right' rally to gain sympathy for the movement after the murder of Heather Heyer by a neo-Nazi in Charlottesville in 2017:

> I'm going to bring a Remington 700 and start shooting Alt-right guys. We need sympathy after that landwhale got all the liberals teary eyed, so someone is going to have to make it look like the left is becoming more violent and radicalized. It's a false flag for sure, but I'll be aiming for the more tanned/dark haired muddied jeans in the crowd so real whites won't have to worry. (Indianapolis Man Arrested for Threatening Boston Free Speech Rally Attendees in 2017 2018)

People in harassment campaigns often embrace a viewpoint that frames them as the oppressed underdog, despite the fact their members are parts of socially powerful dominant groups fighting to preserve the status quo (Lewis 2018, 21–22, 24). They are tempted to invent that oppression when they do not find it.

12. This is also some of the explanation behind why harassment campaigns like Gamergate attack women cited or mentioned by men rather than the men in opposition themselves (Golding and Van Deventer 2016, 178): out of any given online mob, statistically more people will take the time and energy to harass someone outside the presumed 'default' of online culture because of entrenched sexism, racism, homophobia, transphobia, ableism and so on.
13. Or at least, the first *recorded* example of such a guide …
14. Videogame 'walkthroughs' have become more complicated as videogames have themselves grown in complexity; at heart, they combine providing hints and revealing hidden information with suggestions for *how to play the game better*, as an aid to people having trouble negotiating the text. However, it is possible to gain notoriety, respect and fame for having a particularly singular contribution to videogame walkthroughs, as happened for Kao Megura and his guides for *Final Fantasy 7* (Square 1997; Burn and Schott 2004).
15. The infamous 'Tactics for Effective Conservative Blogging' apocryphally attributed to Karl Rove is an example of a guide and one that is still being applied by harassment communities to waste the time of the opposition while pretending to be arguing in good faith (LaCapria 2018).
16. An example of one of Gamergate's 'conspiracy maps' is available here at https://web.archive.org/save/https://imgur.com/zPDtMCK. An example of an image arguing that the videogame *Gone Home* (The Fullbright Company 2013) only received the high review scores that it did due to corruption is available at https://web.archive.org/save/https://imgur.com/sDOjHhM (Jace_Neoreactionary 2014).
17. DARVO stands for 'Deny, Attack, and Reverse Victim and Offender' (Freyd 2016).
18. Discussed in more detail in Chap. 5.

References

/u/coffeeheadphone. 2014. The Everyman's Guide to Gamergate or: How to Have a Real Impact Instead of Posting Angrily on Twitter. *Reddit*. KotakuInAction. 18 September. https://web.archive.org/web/20181023013136/https://www.reddit.com/r/KotakuInAction/comments/2go9a4/the_everymans_guide_to_gamergate_or_how_to_have_a/.

@bethany_lacina. 2018. Updated Data on Twitter before/after @ChuckWendig Being Fired from @Marvel/@StarWars. Tweet. *Twitter* (blog). 18 October. https://twitter.com/bethany_lacina/status/1052973196897595392?ref_src=twsrc%5Etfw.

'weev' Auernheimer, Andrew. 2016. Comments - Feminist Nintendo Employee Alison Rapp the New Voice of Child-Sex Crusaders. *Daily Stormer*. 27 February. https://web.archive.org/web/20160331210419/http://bbs.dailystormer.com/t/feminist-nintendo-employee-alison-rapp-the-new-voice-of-child-sex-crusaders/5096/45.

Alexander, Leigh. 2016. Online Abuse: How Women Are Fighting Back. *The Guardian*. 13 April. https://www.theguardian.com/technology/2016/apr/13/online-abuse-how-women-are-fighting-back.

Alfonso III, Fernando. 2014. "Freebleeding" Is a 4chan Hoax. *The Daily Dot*. 1 February. https://www.dailydot.com/unclick/free-bleeding-is-a-4chan-hoax/.

Allaway, Jennifer. 2014. #Gamergate Trolls Aren't Ethics Crusaders; They're a Hate Group. *Jezebel*. 13 October. http://jezebel.com/gamergate-trolls-arent-ethics-crusaders-theyre-a-hate-1644984010.

Allen, Jay. 2015. How Crowdfunding Helps Haters Profit from Harassment. *Boing Boing*. 14 January. https://boingboing.net/2015/01/14/how-crowdfunding-helps-haters.html.

Ambreen, Sam. 2019. FYI: Kiwi Farms Linked to at least 2 Murders and 4 Suicides. *Left At The Lights* (blog). 8 March. https://web.archive.org/web/20200312021829/https://samambreen.wordpress.com/2019/03/08/fyi-kiwi-farms-linked-to-at-least-2-murders-and-4-suicides/.

Anglin, Andrew. 2016. Feminist Nintendo Employee Alison Rapp Is the New Voice of Child-Sex Crusaders. *Daily Stormer*. 27 February. https://web.archive.org/web/20160331210841/http://www.dailystormer.com/feminist-nintendo-employee-alison-rapp-the-new-voice-of-child-sex-crusaders/.

Asselin, Janelle. 2014. It Happened to Me: I Received Rape Threats After Criticizing a Comic Book'. *XoJane*. 25 April. http://www.xojane.com/it-happened-to-me/janelle-asselin-comic-book-rape-threats.

BBC. 2018a. Toronto Suspect Praised "incel" Killer. *BBC News*, 25 April, sec. US & Canada. https://www.bbc.com/news/world-us-canada-43883052.

———. 2018b. How Rampage Killer Became Misogynist "Hero". *BBC News*, 26 April, sec. US & Canada. https://www.bbc.com/news/world-us-canada-43892189.

Bell, Stewart. 2020. RCMP Adding Incels to Terrorism Awareness Guide. *Global News*. 8 June. https://globalnews.ca/news/7021882/rcmp-incel-terrorism-guide/.

Bernstein, Joseph. 2016. Top Conservative Writer Is a Group Effort, Sources Say. *Buzzfeed*. 1 April. http://www.buzzfeed.com/josephbernstein/top-conservative-writer-is-a-group-effort-sources-say#.dk76VwjJb.

boyd, danah. 2007. Safe Havens for Hate Speech Are Irresponsible. *Apophenia*. 26 March. http://www.zephoria.org/thoughts/archives/2007/03/26/safe_havens_for.html.

Buni, Catherine, and Soraya Chemaly. 2014. The Unsafety Net: How Social Media Turned Against Women. *The Atlantic*. 9 October. https://www.theatlantic.com/technology/archive/2014/10/the-unsafety-net-how-social-media-turned-against-women/381261/.

Burn, Andrew, and Gareth Schott. 2004. Heavy Hero or Digital Dummy? Multimodal Player–Avatar Relations in Final Fantasy 7. *Visual Communication* 3 (2): 213–33. https://doi.org/10.1177/147035704043041.

Butt, Mahli-Ann Rakkomkaew, and Thomas Apperley. 2016. Vivian James - The Politics of #Gamergate's Avatar. In *From the 1st Joint International Conference of DiGRA and FDG*. Dundee, The School of Arts, Media and Computer Games, Abertay University.

Campbell, Colin. 2017. Anita Sarkeesian's Astounding "Garbage Human" Moment. *Polygon*. 27 June. https://www.polygon.com/features/2017/6/27/15880582/anita-sarkeesian-garbage-human-vidcon-interview.

———. 2019. The Anita Sarkeesian Story. *Polygon*. 19 June. https://www.polygon.com/features/2019/6/19/18679678/anita-sarkeesian-feminist-frequency-interview-history-story.

Castello, Jay. 2018. Riot Games Fire at Least One Employee after Diversity Panel Backlash. *Rock, Paper, Shotgun* (blog). 8 September. https://www.rockpapershotgun.com/2018/09/08/riot-games-fires-at-least-one-employee-after-diversity-panel-backlash/.

Cecco, Leyland. 2020. Canada Police Say Machete Killing Was "incel" Terror Attack. *The Guardian*, 19 May, sec. World news. https://www.theguardian.com/world/2020/may/19/toronto-attack-incel-terrorism-canada-police.

Chang, Alvin. 2018. We Analyzed Every QAnon Post on Reddit. Here's Who QAnon Supporters Actually Are. *Vox*. 8 August. https://www.vox.com/2018/8/8/17657800/qanon-reddit-conspiracy-data.

Chess, Shira, and Adrienne Shaw. 2015. A Conspiracy of Fishes, or, How We Learned to Stop Worrying About #GamerGate and Embrace Hegemonic Masculinity. *Journal of Broadcasting & Electronic Media* 59 (1): 208–220. https://doi.org/10.1080/08838151.2014.999917.

———. 2016. We Are All Fishes Now: DiGRA, Feminism, and GamerGate. *Transactions of the Digital Games Research Association* 2 (2) http://todigra.org/index.php/todigra/article/view/39.

Citron, Danielle Keats. 2014. *Hate Crimes in Cyberspace*. Cambridge, MA: Harvard University Press.

Cole, Samantha. 2018. The Complicated Appeal of "Gamer Girl" Porn. *Vice*. 24 September. https://www.vice.com/en_us/article/43835d/gamer-girl-porn-rule-34.

Collins, Ben. 2019. Posing as Gay Men on Twitter, a Troll Goes Viral with Attempts to Falsely Tie the LGBTQ Community to Pedophilia. *NBC News*. 4 January. https://www.nbcnews.com/tech/tech-news/posing-gay-men-twitter-troll-goes-viral-attempts-falsely-tie-n954721.

Cornish, Audie. 2019. *How Gamergate Became A Template For Malicious Action Online*. All Things Considered (NPR). https://www.npr.org/2019/08/30/756034720/how-gamergate-became-a-template-for-malicious-action-online.

Cox, Carolyn. 2014. Female Game Journalists Quit Over Harassment, #GamerGate Harms Women. *The Mary Sue*. 4 September. http://www.themarysue.com/gamergate-harms-women/.

———. 2015. Someone Tried to Photoshop Sikh Games Critic Veerender Jubbal to Look Like a Paris Attacker, Major News Outlets Fell For It. *The Mary Sue*. 16 November. http://www.themarysue.com/veerender-jubbal/.

Cross, Katherine. 2014. What "GamerGate" Reveals About the Silencing of Women. *Rewire*. 9 September. https://rewire.news/article/2014/09/09/gamergate-reveals-silencing-women/.

———. 2015. "Things Have Happened in the Past Week": On Doxing, Swatting, and 8chan. *Feministing*. 16 January. http://feministing.com/2015/01/16/things-have-happened-in-the-past-week-on-doxing-swatting-and-8chan/.

D'Anastasio, Cecilia. 2018a. Inside The Culture of Sexism at Riot Games. *Kotaku Australia*. 8 August. https://www.kotaku.com.au/2018/08/inside-the-culture-of-sexism-at-riot-games/.

———. 2018b. How YouTube Fueled The Anti-Social Justice Movement. *Kotaku*. 20 September. https://kotaku.com/how-youtube-fueled-the-anti-social-justice-movement-1829207455.

D'Souza, Steven. 2018. Bots, Trolls and Fake News: Social Media Is a Minefield for U.S. Midterms. *CBC*. 21 October. https://www.cbc.ca/news/world/national-us-midterm-elections-bots-trolls-fake-news-1.4863258.

Dena, Christy. 2008. Emerging Participatory Culture Practices: Player-Created Tiers in Alternate Reality Games. *Convergence: The International Journal of Research into New Media Technologies* 14 (1): 41–57. https://doi.org/10.1177/1354856507084418.

———. 2009. *Transmedia Practice: Theorising the Practice of Expressing a Fictional World across Distinct Media and Environments*. PhD University of Sydney, Sydney. http://www.christydena.com/phd/.

Desborough, James. 2015. Popular Ludology No.0 – Call for Submissions. *Postmortem Studios*. 30 June. https://postmortemstudios.wordpress.com/2015/06/30/popular-ludology-no-0-call-for-submissions/.

Dewey, Caitlin. 2016. This Horrifying and Newly Trendy Online-Harassment Tactic Is Ruining Careers. *The Washington Post*. 11 April. https://www.washingtonpost.com/news/the-intersect/wp/2016/04/11/this-horrifying-and-newly-trendy-online-harassment-tactic-is-ruining-careers/.

Evon, Dan. 2017. Fact Check: Is "LGBT" Adding a "P" for Pedosexuals? *Snopes.Com*. 7 December. https://www.snopes.com/fact-check/lgbtp-adding-letter/.

Farokhmanesh, Megan. 2018. ArenaNet Firings Cast a Chilling Shadow across the Game Industry. *The Verge*. 12 July. https://www.theverge.com/2018/7/12/17565218/arenanet-guild-wars-firing-games-social-media-harassment.

Feinberg, Ashley. 2017. This Is the Daily Stormer's Playbook. *Huffington Post*, 13 December, sec. Politics. https://www.huffingtonpost.com/entry/daily-stormer-nazi-style-guide_us_5a2ece19e4b0ce3b344492f2.

Fogel, Stefanie. 2018. Video Game Developer Dies After Setting Herself on Fire. *Variety* (blog). 26 June. https://variety.com/2018/gaming/news/chloe-sagal-death-1202858068/.

Frank, Jenn. 2014. On Leaving. *Infinite Lives*. 11 September. http://infinitelives.net/2014/09/11/on-leaving/.

———. 2015. How to Attack a Woman Who Works in Video Gaming. *The Guardian*. 1 September. http://www.theguardian.com/technology/2014/sep/01/how-to-attack-a-woman-who-works-in-video-games.

Freyd, Jennifer J. 1997. Violations of Power, Adaptive Blindness and Betrayal Trauma Theory. *Feminism & Psychology* 7 (1): 22–32.

———. 2016. What Is DARVO? *UOregon.Edu*. http://dynamic.uoregon.edu/jjf/defineDARVO.html.

Futrelle, David. 2014a. Zoe Quinn's Screenshots of 4chan's Dirty Tricks Were Just the Appetizer. Here's the First Course of the Dinner, Directly from the IRC Log. *We Hunted the Mammoth*. 8 September. http://www.wehuntedthemammoth.com/2014/09/08/zoe-quinns-screenshots-of-4chans-dirty-tricks-were-just-the-appetizer-heres-the-first-course-of-the-dinner-directly-from-the-irc-log/.

———. 2014b. Spamming, Doxxing and Sockpuppeting: 4Channers' Dirty Tricks, Straight from Their IRC Log. *We Hunted the Mammoth*. 10 September. http://www.wehuntedthemammoth.com/2014/09/10/spamming-doxxing-and-sockpuppeting-4channers-dirty-tricks-straight-from-their-irc-log/.

———. 2014c. Yet Another Woman in Gaming Has Been Driven from Her Home by Death Threats. *We Hunted The Mammoth* (blog). 11 October. http://www.

wehuntedthemammoth.com/2014/10/11/yet-another-woman-in-gaming-has-been-driven-from-her-home-by-death-threats/.

G2Geek. 2011. Stochastic Terrorism: Triggering the Shooters. *Daily Kos*. 10 January. https://www.dailykos.com/story/2011/1/10/934890/-.

GamerGateOP/Twitter Flooding Instructions. 2014. GitHub. 3 September. https://web.archive.org/web/20140902231940/https://github.com/GamerGateOP/GamerGateOP/blob/master/Twitter%20Flooding%20Instructions.md.

Gardiner, Becky, Mahana Mansfield, Ian Anderson, Josh Holder, Daan Louter, and Monica Ulmanu. 2016. The Dark Side of Guardian Comments. *The Guardian*. 12 April. https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments.

Gardner, Kate. 2018. Bots Probably Helped Get Chuck Wendig Fired, Because Twitter Is a Cesspool. 19 October. https://www.themarysue.com/chuck-wendig-twitter-bots/.

Geiger, R. Stuart. 2016. Bot-Based Collective Blocklists in Twitter: The Counterpublic Moderation of Harassment in a Networked Public Space. *Information, Communication & Society* 19 (6): 787–803. https://doi.org/10.1080/1369118X.2016.1153700.

Golding, Dan, and Leena Van Deventer. 2016. *Game Changers: From Minecraft to Misogyny, the Fight for the Future of Videogames*. South Melbourne: VIC Affirm Press.

Hankes, Keegan. 2018. Evidence of New Mexico School Shooter's Involvement in the Racist "Alt-Right" Is Overwhelming. *Southern Poverty Law Center*. 8 February. https://www.splcenter.org/hatewatch/2018/02/08/evidence-new-mexico-school-shooter%E2%80%99s-involvement-racist-alt-right-overwhelming.

Hern, Alex. 2018. Who Are the "incels" and How Do They Relate to Toronto van Attack? *The Guardian*, 25 April, sec. Technology. https://www.theguardian.com/technology/2018/apr/25/what-is-incel-movement-toronto-van-attack-suspect.

Hess, Amanda. 2014. Why Women Aren't Welcome on the Internet. *Pacific Standard*. 6 January. https://psmag.com/social-justice/women-arent-welcome-internet-72170.

Hirsch, Alex. 2012. Gravity Falls.

Hoverd, Will, Leon Salter, and Kevin Veale. 2020. The Christchurch Call: Insecurity, Democracy and Digital Media. Can It Really Counter Online Hate and Extremism? *SN Social Sciences* Digital Hate and (Anti-)Social Media.

idlediletante. 2015. Sargon of Akkad and Thunderf00t: #Gamergate's Well-Paid Talking Heads. *Daily Kos*. 17 March. https://www.dailykos.com/

story/2015/3/17/1370280/-Sargon-of-Akkad-and-Thunderf00t-Gamergate-s-Well-Paid-Talking-Heads.

Indianapolis Man Arrested for Threatening Boston Free Speech Rally Attendees in 2017. 2018. United States Department of Justice. 8 June. https://www.justice.gov/usao-ma/pr/indianapolis-man-arrested-threatening-boston-free-speech-rally-attendees-2017.

Jace_Neoreactionary. 2014. This Image Explains How Gone Home Got a 10 from Polygon. It Sums up the Problem with the Current Gaming Press/SJW Clique. *KotakuInAction*. 29 September. https://web.archive.org/web/20161007050330/https://www.reddit.com/r/KotakuInAction/comments/2hsv77/this_image_explains_how_gone_home_got_a_10_from/.

Jane, Emma A. 2014. Your a Ugly, Whorish, Slut. *Feminist Media Studies* 14 (4): 531–546. https://doi.org/10.1080/14680777.2012.741073.

Janes, Stephanie. 2019. *Alternate Reality Games: Promotion and Participatory Culture*, Routledge Critical Advertising Studies. Routledge. https://books.google.co.nz/books?id=rFl0ygEACAAJ.

Jaworski, Michelle. 2016. Where Is the Bill Cipher Statue from "Gravity Falls" Located? *The Daily Dot*. 3 August. https://www.dailydot.com/parsec/gravity-falls-bill-statue-located-reedsport-oregon/.

Johnston, Casey. 2014. Chat Logs Show How 4chan Users Created #GamerGate Controversy. *Ars Technica*. 9 September. https://arstechnica.com/gaming/2014/09/new-chat-logs-show-how-4chan-users-pushed-gamergate-into-the-national-spotlight/.

Jong, Carolyn. 2014. "Fighting the Good Fight": GamerGate, NotYourShield, and Neo-Fascism. *Academia.Edu*. December. https://www.academia.edu/10100661/_Fighting_the_Good_Fight_GamerGate_NotYourShield_and_Neo-fascism.

Jubbal, Veerender. 2015. Veerender Jubbal Statement in Response to Fake Paris Terrorism Photo. *The Sikh Coalition*. 16 November. http://sikhcoalition.org/documents/pdf/2015_VeerenderJubbal_Statement.pdf.

Kayyem, Juliette. 2019. There Are No Lone Wolves. *Washington Post*. 4 August. https://www.washingtonpost.com/opinions/2019/08/04/there-are-no-lone-wolves/.

Keats, Jonathon. 2019. How Stochastic Terrorism Lets Bullies Operate in Plain Sight. *Wired*, 21 January. https://www.wired.com/story/jargon-watch-rising-danger-stochastic-terrorism/.

Kim, Jeffrey, Elan Lee, Timothy Thomas, and Caroline Dombrowski. 2009. Storytelling in New Media: The Case of Alternate Reality Games, 2001–2009. *First Monday*, May. https://doi.org/10.5210/fm.v14i6.2484.

Kim, Jeffery Y., Jonathan P. Allen, and Elan Lee. 2010. Alternate Reality Gaming. *Communications of the ACM* 51 (2): 36–42. https://doi.org/10.1145/1314215.1314222.

Klee, Miles. 2014. Creator of "Beat Up Anita Sarkeesian" Says #Gamergate Is Anti-Harassment. *The Daily Dot*. 20 October. http://www.dailydot.com/parsec/creator-beat-up-anita-sarkeesians-says-gamergate-is-anti-harassment/.

Klepeck, Patrick. 2016. The Ugly New Front In The Neverending Video Game Culture War. *Kotaku*. 4 March. http://kotaku.com/the-ugly-new-front-in-the-neverending-video-game-cultur-1762942381.

Kuznekoff, Jeffrey H., and Lindsey M. Rose. 2013. Communication in Multiplayer Gaming: Examining Player Responses to Gender Cues. *New Media & Society* 15 (4): 541–556. https://doi.org/10.1177/1461444812458271.

LaCapria, Kim. 2018. FACT CHECK: Did Karl Rove Write "Tactics for Effective Conservative Blogging"? *Snopes.Com*. 3 May. https://www.snopes.com/fact-check/did-karl-rove-write-tactics-conservative-blogging/.

Lapowsky, Issie. 2018. Here's How Much Bots Drive Conversation During News Events. *Wired*, 30 October. https://www.wired.com/story/new-tool-shows-how-bots-drive-conversation-for-news-events/.

Lewis, Rebecca. 2018. Alternative Influence: Broadcasting the Reactionary Right on Youtube. *Data & Society*. https://datasociety.net/output/alternative-influence/.

lightninggrrl. 2016. I Am Being Stalked and Harassed by Kiwi Farms and SA. *Wrong Planet*. 24 March. http://wrongplanet.net/forums/viewtopic.php?t=308671.

Lum, Zi-Ann. 2015. Veerender Jubbal, Sikh-Canadian Journalist, Wrongly ID'd As Paris Terror Suspect. *Huffington Post*. 16 November. http://www.huffingtonpost.ca/2015/11/16/veerender-jubbal_n_8577520.html.

Lynch, Ashley. 2015. A Final Word on #notyourshield. *Medium*. 24 February. https://medium.com/@ashleylynch/a-final-word-on-notyourshield-628ca5876cec#.4hr3f631v.

———. 2016. Gamers Are Still Over (but They're Not over Trump). *Medium*. 25 October. https://medium.com/@ashleylynch/gamers-are-still-over-but-theyre-not-over-trump-807dde821512?source=user_profile%2D%2D%2D%2D%2D%2D%2D%2D%2D-3-.

Marcotte, Amanda. 2015. The Weird, Woolly and Now Dangerous Fantasy World of Online Misogynists. *Raw Story*. 14 September. http://www.rawstory.com/2015/09/the-weird-woolly-and-now-dangerous-fantasy-world-of-online-misogynists/.

Massanari, Adrienne L. 2015. #Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures. *New Media & Society*. October. https://doi.org/10.1177/1461444815608807.

———. 2018. Rethinking Research Ethics, Power, and the Risk of Visibility in the Era of the "Alt-Right" Gaze. *Social Media + Society* 4 (2) https://doi.org/10.1177/2056305118768302.

Mastroianni, Brian. 2015. After Paris Attacks, Doctored Photo Wrongly Accuses Sikh Man Veerender Jubbal of Being Terrorist. *CBS News*. 16 November. http://www.cbsnews.com/news/sikh-man-wrongly-accused-of-being-paris-terrorist-in-altered-image/.

McGonigal, Jane. 2003a. A Real Little Game: The Performance of Belief in Pervasive Play. http://www.avantgame.com/MCGONIGAL%20A%20Real%20Little%20Game%20DiGRA%202003.pdf.

———. 2003b. This Is Not a Game: Immersive Aesthetics and Collective Play. http://www.seanstewart.org/beast/mcgonigal/notagame/paper.pdf.

———. 2010. The Puppet Master Problem: Design for Real-World, Mission Based Gaming. In *Second Person: Role-Playing and Story in Games and Playable Media*, ed. Pat Harrigan and Noah Wardrip-Fruin, 251–263. Cambridge, MA; London: MIT Press. http://www.avantgame.com/McGonigal_THE-PUPPET-MASTER-PROBLEM_MITpress.pdf.

McGuill, Dan. 2018. Fact Check: Did Nike Offer "People of Color" a 75 Percent Off Coupon? *Snopes.Com*. 7 September. https://www.snopes.com/fact-check/nike-coupon-color/.

McKibben, Bill. 2016. Embarrassing Photos of Me, Thanks to My Right-Wing Stalkers. *The New York Times*. 5 August. http://www.nytimes.com/2016/08/07/opinion/sunday/embarrassing-photos-of-me-thanks-to-my-right-wing-stalkers.html?smid=tw-share&_r=1.

McWhertor, Michael. 2014. Game Developer Brianna Wu Flees Home after Death Threats, Mass. Police Investigating. *Polygon*. 11 October. https://www.polygon.com/2014/10/11/6963279/brianna-wu-death-threats-police-harassment.

Meer, Alec. 2010a. Is This Portal 2? *Rock, Paper, Shotgun*. March. http://www.rockpapershotgun.com/2010/03/02/is-this-portal-2/.

———. 2010b. Portal: There's Something Going On. *Rock, Paper, Shotgun*. March. http://www.rockpapershotgun.com/2010/03/02/portal-theres-something-going-on/.

———. 2010c. Valve Won't Let Us Sleep: Portal Updates. *Rock, Paper, Shotgun*. March. http://www.rockpapershotgun.com/2010/03/03/valve-wont-let-us-sleep-portal-updates/.

Megarry, Jessica. 2014. Online Incivility or Sexual Harassment? Conceptualising Women's Experiences in the Digital Age. *Women's Studies International Forum* 47 (November): 46–55. https://doi.org/10.1016/j.wsif.2014.07.012.

Meifert-Menhard, Felicitas. 2013. *Playing the Text, Performing the Future: Future Narratives in Print and Digiture*. Berlin: De Gruyter. http://ezproxy.massey.ac.nz/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=661681&site=eds-live&scope=site.

Mortensen, Torill Elvira. 2016. Anger, Fear, and Games: The Long Event of #GamerGate. *Games and Culture*. April. https://doi.org/10.1177/1555412016640408.

Murdock, Sebastian. 2018. California Wildfire Suspect Posted About QAnon, Other Conspiracies. *Huffington Post*. 10 August, sec. Crime. https://www.huffingtonpost.com/entry/california-wildfire-suspect-posted-about-qanon-other-conspiracies_us_5b6dc69de4b0ae32af97e953.

Neiwert, David. 2015. Illinois Woman With Neo-Nazi Leanings Charged in Canadian Mass Murder Plot. *Southern Poverty Law Center*. 18 February. https://www.splcenter.org/hatewatch/2015/02/18/illinois-woman-neo-nazi-leanings-charged-canadian-mass-murder-plot.

O'Connor, Alice. 2018. Riot Games Apologise for Workplace Culture They Let Fester, Vow to "Become a Leader on Diversity, Inclusion, and Culture". *Rock, Paper, Shotgun* (blog). 31 August. https://www.rockpapershotgun.com/2018/08/31/riot-games-apologise-for-workplace-culture/.

———. 2019. New Zealand Banned a Mass Shooting Game as a "Terrorist Publication". *Rock, Paper, Shotgun* (blog). 1 November. https://www.rockpapershotgun.com/2019/11/01/new-zealand-banned-a-mass-shooting-game-as-a-terrorist-publication/.

Phillips, Whitney. 2015. *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture*. Cambridge, MA: MIT Press.

Pless, Margaret. 2014. Eron Gjoni, Hateful Boyfriend. *Internet Famous Angry Men*. 6 December. http://idledillettante.com/2014/12/06/eron-gjoni-hateful-boyfriend/.

———. 2016. Kiwi Farms, the Web's Biggest Stalker Community. *New York Magazine*. 19 July. http://nymag.com/selectall/2016/07/kiwi-farms-the-webs-biggest-community-of-stalkers.html.

Raymond, Eric. 2016. Name Them and Shame Them. *Google Plus*. 11 April. https://plus.google.com/+EricRaymond/posts/Wp1ycW9MZrV.

Robb, Amanda. 2017. Pizzagate: Anatomy of a Fake News Scandal. *Rolling Stone* (blog). 16 November. https://www.rollingstone.com/politics/politics-news/anatomy-of-a-fake-news-scandal-125877/.

Robertson, Adi. 2014. Trolls Drive Anita Sarkeesian out of Her House to Prove Misogyny Doesn't Exist. *The Verge*. 27 August. https://www.theverge.com/2014/8/27/6075179/anita-sarkeesian-says-she-was-driven-out-of-house-by-threats.

Romano, Aja. 2014. 4chan Continues Battle against Zoe Quinn, Hacks and Doxes Phil Fish. *The Daily Dot*. 22 August. https://www.dailydot.com/parsec/4chan-hacks-phil-fish-over-his-defense-of-zoe-quinn/.

Rowe, Don. 2019. The Online Cesspits Where Hate Found a Home. *The Spinoff* (blog). 19 March. https://thespinoff.co.nz/media/19-03-2019/the-online-cesspits-where-hate-found-a-home/.

Sarkeesian, Anita. 2012. Anita Sarkeesian at TEDxWomen 2012. *TEDxTalks*. 5 December. https://www.youtube.com/watch?v=GZAxwsg9J9Q.

———. 2014a. I Usually Don't Share the Really Scary Stuff. But It's Important for Folks to Know How Bad It Gets [Trigger Warning]. Tweet. *@femfreq* (blog). 27 August. https://twitter.com/femfreq/status/504718160902492160/photo/1.

———. 2014b. Anita Sarkeesian Speaking at XOXO Conference. *Feminist Frequency*. 7 October. https://feministfrequency.com/video/anita-sarkeesian-speaking-at-xoxo-conference/.

———. 2015a. One Week of Harassment on Twitter. *Feminist Frequency*. 27 January. https://feministfrequency.com/2015/01/27/one-week-of-harassment-on-twitter/.

———. 2015b. Harassment Through Impersonation: The Creation of a Cyber Mob. *Feminist Frequency*. 10 December. http://feministfrequency.com/2015/12/10/harassment-through-impersonation-the-creation-of-a-cyber-mob/.

———. 2016. On Twitter, Conspiracy Theories, and Information Cascades. *Feminist Frequency*. 22 February. https://feministfrequency.com/2016/02/22/on-twitter-conspiracy-theories-and-information-cascades/.

Schnier, Michael. 2014. Brand Management After Operation Disrespectful Nod: A Brave New World. *Medium*. 29 October. https://medium.com/the-internet-made-me-do-it/brand-management-after-operation-disrespectful-nod-a-brave-new-world-59a393501c26#.hbo0wc6yh.

Schubert, Damion. 2015. The Terrifying, Pathetic Case of Joshua Goldberg. *Zen of Design*. 13 September. http://www.zenofdesign.com/the-terrifying-pathetic-case-of-joshua-goldberg/.

Shaw, Frances. 2013. Still "Searching for Safety Online": Collective Strategies and Discursive Resistance to Trolling and Harassment in a Feminist Network. *The Fibreculture Journal*, no. 22. http://twentytwo.fibreculturejournal.org/fcj-157-still-searching-for-safety-online-collective-strategies-and-discursive-resistance-to-trolling-and-harassment-in-a-feminist-network/.

Sheldon, Lee. 2010. Ilovebees: Playing and Designing in Real-Time. In *Well Played 2.0: Video Games, Value and Meaning*, ed. Drew Davidson. ETC Press.

Shepherd, Tamara, Alison Harvey, Tim Jordan, Sam Srauy, and Kate Miltner. 2015. Histories of Hating. *Social Media + Society* 1 (2) https://doi.org/10.1177/2056305115603997.

Sierra, Kathy. 2014. Trouble at the Koolaid Point. *Serious Pony*. 7 October. http://seriouspony.com/trouble-at-the-koolaid-point.

Sinders, Caroline. 2015. That Time the Internet Sent a SWAT Team to My Mom's House. *Narrative.Ly*. 17 July. http://narrative.ly/that-time-the-internet-sent-a-swat-team-to-my-moms-house/.

Smith, Graham. 2018. ArenaNet Throw Two Guild Wars 2 Writers to the Wolves. *Rock, Paper, Shotgun* (blog). 6 July. https://www.rockpapershotgun.com/2018/07/06/arenanet-throw-two-guild-wars-2-writers-to-the-wolves/.

Social Justice Viv. 2016. Kiwi Farms, the Web's Biggest Community of…. *Tumblr*. 13 September. http://socialjusticeviv.tumblr.com/post/150364975017/kiwi-farms-the-webs-biggest-community-of.

Square. 1997. *Final Fantasy VII*. Square Product Development Dept. #1

Stanton, Rich. 2015. Gamergate Supporters Are Responsible for the Terrorist Photoshopping of Journalist Veerender Jubbal. *Vice*. 17 November. https://www.vice.com/en_uk/read/gamergate-members-are-responsible-for-the-terrorist-photograph-of-journalist-veerender-jubbal-503.

Stewart, Keith. 2014. Zoe Quinn: "All Gamergate Has Done Is Ruin People's Lives". *The Guardian*. 3 December. https://www.theguardian.com/technology/2014/dec/03/zoe-quinn-gamergate-interview.

Stewart, Sean, Elan Lee, and Jordan Weisman. 2001. The Beast.

Stewart, Sean, Elan Lee, and Jane McGonigal. 2004. I Love Bees. http://www.ilovebees.com.

Stuart, Keith. 2014. Brianna Wu and the Human Cost of Gamergate: "Every Woman I Know in the Industry Is Scared". *The Guardian*, 17 October, sec. Games. https://www.theguardian.com/technology/2014/oct/17/brianna-wu-gamergate-human-cost.

Szulborski, Dave. 2005. *Through the Rabbit Hole: A Beginner's Guide to Playing Alternate Reality Games*. LuLu.com.

Tait, Maggie. 2019. Two Terrorist Publications Banned. *NZ Office of Film & Literature Classification*. 31 October. https://www.classificationoffice.govt.nz/news/latest-news/two-terrorist-publications-banned.

Taub, Amanda. 2016. The Guardian Study's Hidden Lesson: Trolls Reinforce White Male Dominance in Journalism. *Vox*. 13 April. http://www.vox.com/2016/4/13/11414942/guardian-study-harassment.

The Fullbright Company. 2013. *Gone Home*. Portland, OR: The Fullbright Company.

Valve. 2007. *Portal*. Valve Corporation.

Veale, Kevin. 2012. *Comparing Stories: How Textual Structure Shapes Affective Experience in New Media*. Auckland: University of Auckland. http://hdl.handle.net/2292/10347.

———. 2013. Capital, Dialogue, and Community Engagement—'My Little Pony: Friendship Is Magic' Understood as an Alternate Reality Game. *Transformative Works and Cultures* 14 (September) https://doi.org/10.3983/twc.2013.0510.

Vossen, Emma. 2018. On the Cultural Inaccessibility of Gaming: Invading, Creating, and Reclaiming the Cultural Clubhouse. *UWSpace*. http://hdl.handle.net/10012/13649.

Walschots, Natalie. 2015. Gamergate: The Greatest Trick The Devil Ever Pulled. *The Establishment*. 18 November. https://medium.com/the-establishment/gamergate-the-greatest-trick-the-devil-ever-pulled-876aa73e3d2e.

Zadrozny, Brandy, and Ben Collins. 2018. Who Is behind the Qanon Conspiracy? We've Traced It to Three People. *NBC News*. 14 August. https://www.nbcnews.com/tech/tech-news/how-three-conspiracy-theorists-took-q-sparked-qanon-n900531.

CHAPTER 4

Gaming the Rules

The focus of the last chapter was in explaining the ways that online harassment campaigns function as alternate reality games (ARGs), and the focus of this chapter is to explore one such connection in more detail. Specifically, this chapter will explore the ways that members of both ARGs and harassment communities learn both social and technological 'rules' in order to manipulate them to achieve their goals.

It will explore a series of case studies, each of which illustrates how harassment communities reverse-engineer the designs of networked publics in order to discover how to turn them to their advantage. These case studies will illustrate, among other things, that some popular approaches pitched as solutions to online harassment are distractions that are very unlikely to work, and are likely to empower harassment communities further.

As discussed in the last chapter, ARGs illustrate the fearsome power that collective-intelligences have to solve problems: an ARG community solved three months' worth of content for *The Beast* in one day (Kim et al. 2009; McGonigal 2003b, 2–3; Meifert-Menhard 2013, 161; Sheldon 2010, 277). Some of their marked achievements are technological, such as where a community created and distributed software to complete the missing part of a long password so quickly that the puppet masters did not notice the problem until it was already resolved (McGonigal 2003a, 12–13). However, alongside technological achievements, they also solve problems through learning and gaming social rules—often in ways the

K. Veale, *Gaming the Dynamics of Online Harassment*,
https://doi.org/10.1007/978-3-030-60410-3_4

people running the game do not expect. For example, Jane McGonigal discovered players praising the acting of a 'plant' they discovered at a hotel, who they had persuaded to allow them access to employee-only areas of the hotel in order to achieve their goals (McGonigal 2003a, 20). What horrified McGonigal was that the ARG had no such plant, and the players had effectively broken into a hotel through social engineering without realising that was what they were doing.

We can see the same ability to deploy both technological and social skills in harassment communities, and that is part of the problem. One example I personally witnessed was on 4chan,[1] where members of Anonymous hunted down someone who was posting videos where they abused animals to the site. There were many technological dimensions to the hunt, such as finding embedded location information in the pictures and video to narrow down the location. However, once the location had been narrowed down to a street through a variety of other means, someone solved the rest of the problem with social engineering. They dressed as a utility worker and talked their way into houses 'looking for a leak' in order to gain access and look for the rooms in the backgrounds of the videos. Having located the building through these methods, Anonymous simultaneously reported the abuser to the police and targeted them and their family for harassment.

Harassment communities apply the same distributed-intelligence problem-solving typical of ARGs to finding ways that the underlying affordances and community dynamics of networked publics can be turned into tools both for abuse, and for overcoming any defences presented by their targets (Lomas 2014). For example, Gamergate's use of Twitter was partly motivated by the fact users cannot delete responses to their posts within 'their' feed, meaning they cannot easily get away from abuse (Salter 2017, 254). Essentially, harassment communities are well aware of the ways that networked publics are shaped by the technological affordances that provide the substrates they grow on. Since any given harassment community will grow across multiple different networked publics in pursuit of the people it wishes to terrorise, its members seek to learn how to turn each context to their collective advantage, and the results can be devastating.

Algorithms

Algorithms designed to help with online searches or recommendations are easily weaponised tools for the purposes of abuse and have a disproportionately vast impact because of how much society relies on them to be 'neutral.' 'Google bombs' are where a harassment campaign organises to create as much material that repeatedly links key phrases together as they can, such as the name of someone they want to target and a crime to frame them for (Citron 2014, 69–72). If enough articles are posted claiming that someone is a terrorist, then seemingly neutral web-searches for that person's name will increasingly return articles claiming that they are a terrorist. Gamergate used this technique frequently, such as where the harassment community framed Sarah Nyberg, Alison Rapp and Dan Olson as paedophiles,[2] Randi Harper as an animal abuser and Veerender Jubbal as a terrorist (Lum 2015; Olson 2014; Pless 2015; Wilson 2016). Having set up this kind of association in online searches, the next step is to contact the workplaces of the people the community is targeting in order to demand to know how the company can justify employing a depraved criminal. If the employer does a web-search to verify these claims—claims magnified greatly via repetition through the harassment community and its guides—then it will look like the accusations have substance. Alison Rapp was fired by Nintendo of America after these tactics were used against her, alongside many other less high-profile cases who had their livelihoods threatened or destroyed.[3] Once a Google bomb is established, at that point the harassment community has outsourced its campaign of terror to Google, massively magnifying its reach and impact in the process.

YouTube's recommendation algorithms serve a similar function, but can be delegated the job of abusing the targets of harassment communities even more easily. The recommendation engines that suggest videos related to those someone has already watched mean that it is now functionally difficult to watch material tied to videogame culture without encountering harassment videos, due to the volume of harassment content linked to the topic (Harper 2016). People who have been targeted by harassment storms have spoken anonymously about the impact this has on their ability to use a ubiquitous pillar of internet culture: they cannot post a video of their cat without knowing that anyone who watches it will likely be recommended videos tied to their name, where members of the harassment community slander and/or abuse them. They are effectively barred from

using YouTube unless they want to contribute to amplifying the voices of those seeking to terrorise them, and helping them profit from their abuse.[4]

James Bridle describes YouTube's algorithms as 'infrastructural violence' in the context of how they drive the creation and recommendation of disturbing content to children in a wider pursuit of profit, and arguably, this is a different manifestation of the same dynamics (Bridle 2017, 2018; Hern 2017). Rebecca Lewis has done extensive work illustrating that YouTube's algorithms form the underlying spine of what she calls an 'alternative influence' network of neo-Nazis, white-supremacists, anti-feminists and other extremist groups online (D'Anastasio 2018; Lewis 2018). YouTube's algorithms mean members of the audience that enter the network from any direction will be introduced to more and more extreme content through the recommendation engine. As a result, YouTube's algorithms both provide a means of interconnection between those creating content in the network and encourage the creation of more extreme content, because it will be rewarded both financially and with more views (Lewis 2018, 36–42). Since the algorithms encourage the creation of more extreme content and distribute it to widen their audience, they directly perform a role in expanding the tiers of extremist and harassment communities. Facebook's algorithms follow the same dynamics, and this was known to Facebook for years while executives scuttled attempts to fix the problem and kept the information from the public:

> The high number of extremist groups was concerning, the presentation says. Worse was Facebook's realization that its algorithms were responsible for their growth. The 2016 presentation states that "64% of all extremist group joins are due to our recommendation tools" and that most of the activity came from the platform's 'Groups You Should Join' and 'Discover' algorithms: "Our recommendation systems grow the problem." (Horwitz and Seetharaman 2020)

As covered in Chap. 3, growing the 'audience' tier of a harassment community also grows the most active tiers. As the size of the most active tiers grows, the odds of someone in the harassment community being willing to perform attacks in physical spaces rises. As a result, the current algorithmic business models applied by both YouTube and Facebook increase the chances of future terrorist attacks.

In some cases, algorithms can become tools for abuse and financial gain simultaneously: content on YouTube tied to LGBTQA+ issues,

particularly anything related to support for the trans community, has been systematically demonetised at the same time as it has been tied to ads for an anti-LGBTQA+ organisation that the Southern Poverty Law Centre categorises as a hate group (Alexander 2018; Bardo 2018; Farokhmanesh 2018; 'Nerd City' 2019).[5]

As will be discussed in more detail in Chap. 5, technology industries are very fond of proposing that algorithmic tools are a solution to online harassment. However, given the ease with which harassment communities turn algorithms against their targets, it is very likely that these initiatives will simply provide more weapons for the enemy.

Reporting and Blocking Tools

Infrastructure for reporting posts on social media forms one of the most effective tools in the arsenal of harassment communities because of the evidence that they respond to the volume of reports and not the content of the posts reported. For example, many people have catalogued online abuse on platforms such as Twitter, often including direct and specific threats of rape, violence and murder, alongside the messages from Twitter declaring that they are not against Twitter's rules or terms of service (Buni and Chemaly 2014; Hudson 2015; Sarkeesian 2015; West 2014).[6] Some people have had accounts suspended for resharing content threatening them, while those who posted the threatening material are left alone ('@BechdelCast' 2018; Golding and Van Deventer 2016, 91; Masnick 2017). Given the lack of transparency behind these systems, we are left to speculate as to what the underlying logic behind these decisions is. The central observable pattern seems to be that posts with numbers of reports above a given-but-unknown threshold will be acted on and the accounts suspended, whereas complaints about posts with report numbers below that threshold will be told that the content is not against the rules, regardless of what the content is (Lomas 2017).

Harassment communities find reporting tools fruitful options to experiment with precisely because they can take away someone's control of their account and/or posts. The Kiwi Farms harassment community is a particularly horrific example. They use reporting tools on platforms like Facebook or YouTube to mass-report any mentions of suicidal thoughts or activity by the people they target, in order to have those posts restricted or erased. The explicit goal is to limit the number of people who see the posts and thus limit the possibility that the people terrorised by Kiwi

Farms will receive help, as part of the community's goal of driving people to suicide—a goal they have been successful at on more than one occasion (Ambreen 2019; Fogel 2018; 'lightninggrrl' 2016; 'Social Justice Viv' 2016; Pless 2016).

Harassment communities are not restricted by the codified rules presented by social network platforms, either: they are equally adept at taking legislative changes and weaponising them via mass-reporting in service of abuse. For example, the FOSTA/SESTA[7] law changes targeting sex-workers under the nominal fig-leaf of restricting human trafficking in 2018 have been used to attack women online. In this case, the method has been to mass-report their accounts, claiming that they are sex-workers ('violetblue' 2018), in order to get them dropped by PayPal and/or Patreon.

Another example came after Mandy Morbid accused Zak Smith[8] of rape and abuse (Nagy 2019), inspiring others to come forward with their own experiences (Grey 2019)—experiences which a long-term defender of Smith[9] considered credible ('pjamesstuart' 2019). Smith then sued a number of people—including Morbid—in civil court for defamation (Smith 2019), in jurisdictions where the burden of proof is placed onto the person accused, effectively making the suit easier ('Tenkar' 2019). This has forced Morbid to seek crowd-funding to defend the case (Girdwood 2019), and others have been required to make public apologies as part of out-of-court settlements. The legally mandated apology was posted in a number of online spaces where the 'defamatory' statements were never made, but where Smith had previously been banned from (RPG.net 2013). One possible conclusion is that the goal is to achieve out-of-court settlements by outlasting the defence's finances in jurisdictions with lowered protections for speech, in order to use the apologies as tools in ongoing conflicts online.

Blocking tools on social media platforms are intended so that users have some level of control over who they engage with using the service, but are trivially easy to circumvent: in order to evade a block, the affected harasser need only log out and manually return to the account that has blocked them. At that point, they can read every post or tweet in complete impunity, potentially feeding them (or screenshots of them) back to other members of the harassment community. Another weakness to the system is that there is often nothing to prevent someone blocked from simply creating a new 'sockpuppet' account with which to continue the harassment, since the block will only apply to the original account and needs to

be reapplied manually. This creates an exhausting dynamic where it is more effort to block tides of new harassing accounts as they surface than it is to create them.

In general, the design underlying the dynamics of networked publics means users have little control over their personal experiences in the space, and this lack of control is turned against them by harassment campaigns. They look for avenues of attack that are more energy to respond to or attempt to thwart than they are to cause harm with and grind their targets down—effectively aided by the platforms they operate within.

Context Collapse

Another scenario where harassment communities can cause harm more easily than their targets can prevent it involves turning peoples' statements against them by weaponising a process that social media platforms begin by themselves. Social media innately causes a level of 'context collapse' by bringing different parts of our lives into one shared platform. This limits our ability to present different sides of ourselves in distinct circumstances—such as how we present ourselves differently to family than we do to friends or co-workers (boyd 2011; Chess and Shaw 2015, 2016; Golding and Van Deventer 2016, 97; Litt and Hargittai 2016; Marwick and boyd 2011; Trepte 2015; van der Nagel and Frith 2015). Harassment communities are adept at combing through potentially years' worth of material posted by someone they seek to terrorise, looking for anything that would be damaging if taken out of its original context (Massanari 2018, 3–4).[10] These posts can then be shared to a much broader audience than was originally intended in order to attract negative attention and 'justify' mass-reporting attacks on the author. For example, Katie Notopoulos has written about her experiences of having her Twitter account locked in 2017. An ironic Twitter response to a friend in 2011 that read 'kill all white people,' was mass-reported as a racist threat as part of harassment community retaliation against random targets after Twitter de-verified a number of prominent white-supremacist/neo-Nazi accounts (Notopoulos 2017).

Different social media platforms have affordances that can potentially cause hostile context collapses at an automated, algorithmic level. One example is that it is not possible to delete Tumblr posts that have been reshared, even if the account that created them is entirely deleted: the infrastructure of the site guarantees that authors lose control of their posts

if they are shared, no matter the circumstances ('Coyote' 2019; Williams n.d.). Tumblr makes this situation worse, since its design means that the only forms of interaction that are threaded and have a continuity of conversation are reshares, which encourages resharing. This single piece of design has a huge impact on the networked public which grows within the platform:

> The Tumblr reblog-addition system fosters, firstly, A) reblogging to disagree. The problem here isn't that people get into fights, of course; the problem is that when you reblog to disagree, disputing an objectionable post involves *spreading that post in the process*.
>
> Consequently, a site culture of reblogging-to-disagree means B) more people sharing the worst posts on the site in order to disagree with them. This has the effect of C) stressing you out from seeing all the bad posts on your dash (especially in those long unpredictable chains).
>
> When people are stressed out, they become less patient with each other and quicker to anger. From here, the userbase D) develops a culture of mockery, passing some posts around just to make fun of them, E) lends more visibility to aggression and hostility than to nuance or apologies, and F) makes major fights become inescapable. In these ways, the reblog-addition system facilitates an overall stressful, hostile atmosphere, which undermines community. ('Coyote' 2019)

Effectively, any post someone reshares automatically experiences context collapse, at the same time as the author loses any control over them. 'Coyote' notes that these problems are substantially worsened since it is not possible to turn off reshares and since no corrections made by the author will appear in reshared versions. As a result, there is no ability to correct mistakes, which combined with the inevitability of context collapse produces a deeply unforgiving space for users:

> This can leave Tumblr users open to endless criticism: As more and more people discover the uncorrected version of a post, more and more will try to correct it, unaware that the poster has already got the memo. In severe cases, the inability to stem the tide is what sometimes has made it necessary for Tumblr users to completely delete their blogs in order to escape. Incapacity for retraction makes people unnecessarily vulnerable, which undermines community. ('Coyote' 2019)

Tumblr's design effectively encourages behaviour that duplicates many of the patterns of harassment communities all by itself, which produces a networked public many people find innately hostile and unpleasant.

Further examples of context collapse being part of a platform's foundational design can be found on Twitter and Reddit. Different subreddits form their own community norms and practices, and 'trending' posts or subreddits become visible to a vastly wider audience on Reddit's front page. As a result, increased visibility often produces an increase in harassment community activity targeted against marginalised groups every time a post or subreddit from their community is lifted over the parapet by Reddit's algorithmic recommendation engines (Massanari 2015, 10). Twitter hashtags likewise become visible to a very broad audience once they begin 'trending,' which can often lead to harassment.[11] An example unfolded in 2015 when the DiGRA Australia conference used the #digraa hashtag and produced enough engagement that it eventually became visible as a trending topic. Gamergate had been unaware of the hashtag until that point and immediately flooded it once Twitter raised it to visibility because they were already targeting DiGRA members.[12]

Hashtags

Hashtags provide a very visible microcosm in which to explore the ways that harassment communities learn how they can use the affordances presented by different social media platforms as weapons or to communicate internally within the community.[13] Casey O'Donnell and Mia Consalvo have written about the ways that the Twitter app TweetDeck shaped how Gamergate used Twitter, by making it possible to compartmentalise Twitter feeds around particular hashtags (O'Donnell and Consalvo 2015).[14] With TweetDeck, users could create columns of tweets solely dedicated to the #gamergate hashtag, making it easier for the members of the harassment community to have a shared space within the platform. It was almost exclusively used by the pro-Gamergate cause, and barely ever by anyone critiquing it—partly because anyone critiquing the hashtag came under immediate, sustained abuse (Baio 2014; Salter 2017, 254). It also allowed for dynamics where members of the harassment community would use a four-step process to highlight new targets while denying responsibility for the resulting abuse:

1. Search for potential targets who are discussing Gamergate or otherwise connected to people targeted by the harassment campaign.
2. Engage a target on Twitter 'politely,' using the #Gamergate hashtag.[15]
3. If they responded—and potentially if they did not—'tag in' the Twitter IDs of other members of the harassment community to the conversation, flagging the target for them.
4. Start a dogpile of abuse alongside more polite engagement, at the same time as dropping the #gamergate hashtag from all harassing posts.

This informal process allowed members of the community to simultaneously highlight targets for harassment while claiming that since the abusive posts (mostly)[16] did not contain the #Gamergate hashtag, members of the community were not to blame. It was also a way of shaping what uninvolved people searching for 'gamergate' rather than the hashtag itself would encounter. They were as likely to see the people being targeted by the community describing Gamergate in vitriolic terms because of the abuse they were going through as they were to see any abuse themselves. As a result, they might potentially conclude that 'both sides' were as bad as each other, playing into the hands of the harassment community.

Search Tools

The infrastructure of search tools, and how those tools function on different platforms, is also something harassment communities learn—both for targeting crowdsourced terrorism and to conceal the harassment community's own activities from outside scrutiny. As discussed earlier, one of the fastest methods to check for abuse in a networked public is to start talking about it, because doing so often summons harassers to deny that harassment is happening. The Gamergate harassment community was so active in searching for anyone using 'gamergate' as a term in order to harass their critics that people sought colloquial alternatives, such as 'gators,' 'GG,' 'goobergate' and others to discuss events without attracting abuse. Natalie Walschots was dissatisfied when a conference of the Canadian Games Studies Association asked attendees not to mention Gamergate in any tweets about the conference to avoid the same kind of flood that hit DiGRA Australia in 2015, since it interfered with the ability of the people at the conference to do their work. Her solution was to replace

#gamergate with #deatheaters instead, referring to the magical neo-Nazis from the *Harry Potter* storyworld (Goodyear 2015). Walschots' solution worked until a single person lapsed by mentioning Gamergate at the end of the conference, at which point the dogpile descended on the conference and the harassment campaign began organising against her.

A flipside where harassment communities have developed strategies to avoid searches can be found in 'echoes,' where neo-Nazis created symbolism that would be visible to online searches by the harassment community, but unthreatening to people from outside of it:

> Neo-Nazis, anti-Semites and white nationalists have begun using three sets of parentheses encasing a Jewish surname—for instance, (((Fleishman)))—to identify and target Jews for harassment on blogs and major social media sites like Twitter. As one white supremacist tweeted, "It's closed captioning for the Jew-blind." (Fleishman and Smith 2016)

Just as with any coded communication, as soon as 'echoes' were no longer secret, they stopped performing the function they had been intended for within the harassment community. The same is true of other codes, such as 'Operation Google'—an initiative begun on 4chan to replace slurs for different marginalised groups with terms commonly found on social media, like Google, Skype or Skittles (Ehrenkranz 2016). The central problem is that although codes like these seem relatively harmless through being easy to decode and explain once they are publicised, they are simple and effective for harassment campaigns until they *are* publicised. Additionally, they require no specialist knowledge or skill to generate: people can just make them up.

Chapter 3 explored how harassment campaigns demonstrate the same signs and behaviours as ARGs, meaning we can imagine them as feral and autonomous ARGs whose goal is causing harm. Chapter 4 extended this line of reasoning to explore specific ways that harassment campaigns demonstrate the same ferocious ability to learn and then manipulate both social and technological rules as part of achieving their goals. The next two chapters build on this knowledge by examining ways that treating harassment campaigns as ARGs provides insights that can be used to limit their impact in future. If we have a collective body of knowledge about how to make ARGs work, then by implication we can apply that knowledge to understanding how to stop them from working, or at least blunt their ability to work.

However, Chap. 5 will illustrate that one of the challenges such an initiative would need to overcome is the fact that technology companies are doubtless aware of all of these dynamics covered so far, and yet have done nothing.

Notes

1. I have not been able to find external verification of the details in this specific case. For a related case, see (O'Brien 2009).
2. In an example of the deep hypocrisy that harassment communities often embrace, Dan Olson was accused of paedophilia based on the fact he published an article proving that 8chan—a forum that hosted Gamergate after it was banned from 4chan—was hosting child pornography. Olson's article included evidence in the form of ethically sanitised screenshots from 8chan illustrating its child pornography, and on that basis, Gamergate argued he was distributing child porn.
3. Just as with the firing of Jessica Price and Peter Fries, harassment communities depend on the cowardice and lack of support provided by companies across the creative and technology industries in order to achieve this kind of impact. Lana Polansky argues that the situation goes beyond cowardice into active collaboration: that videogame and technology industries find the existence of harassment communities to be convenient, since it helps keep their workforce 'afraid, quiet, and deprived of leverage' (Polansky 2018).
4. It is also incredibly demotivational to know that if you upload a more critical video, the harassment community will use it as fodder for their own response videos—and probably make significantly more money from the harassment community through doing so than you will.
5. It is worth highlighting that as the Nerd City group have argued, the underlying business model of YouTube's algorithms contributes to the marginalisation of vulnerable groups even outside of situations where it is gamed for abuse ('Nerd City' 2019). They have shown that including particular keywords will result in a video losing any ability to monetise itself via YouTube, despite YouTube's claims to the contrary, and the 'learning algorithm' underlying the system now automatically demonetises LGBTQA+ content. This is another example where harassment campaigns have definitely gamed systems, but large corporations and their algorithmic systems have effectively taken on the job of inflicting that abuse themselves.
6. Twitter even concluded that threats made by the person ultimately responsible for the mailbombing campaign targeting prominent Democrats and their associates in October of 2018 were not against Twitter's rules—until

bad publicity around his arrest for domestic terrorism made them re-evaluate the decision (Beschizza 2018; '@ryanjreilly' 2018).

7. hese are both American laws: a House bill known as FOSTA, the Fight Online Sex Trafficking Act, and the Senate bill, SESTA, the Stop Enabling Sex Traffickers Act.
8. Aka Zak Sabbath, Zak S. and others.
9. Smith has been accused of harassment on many occasions—by different people, across different contexts and across many years—some of whom documented their experiences with him in great detail ('ArthurTheRef' 2019; Hatfield and Ellison 2014; Hurley 2014a, 2014b, 2014c; Kreider 2014, 2015; Matijevic 2015; Tabletop's Missing Stairs 2019). Evidence is available online of Smith informing his followers about particular targets next to statements such as 'Destroy' or 'Get at him,' risking stochastic terrorism ('@FreyjaErlings' 2015; Smith 2015a, 2015b). Other accounts allege that Smith avoids direct harassment of targets when it can be delegated to his followers, creating plausible deniability through tiering effects (Hill 2017). There is evidence he has created 'sockpuppet' accounts to impersonate people online, visible when he posted a statement in the wrong account before deleting it and reposting in the correct one (Author Unknown 2017). The allegations of rape and abuse also accuse him of impersonating people such as Morbid online, using her accounts to defend himself in her name (Nagy 2019).
10. This is an example of asymmetrical information gathering, where it is easier for the community to gather information about their target than it is for the target to discover who is attacking them.
11. Promotional bot accounts on Twitter which mindlessly reshare any mention of a brand or product name as publicity often cause similar problems: anyone critical of the product or brand will have their comment shared with its most rabid fans.
12. Searching for #digraa now reveals a core-sample of posts from the 2015 conference, both during and before Gamergate began flooding it and attacking the people using it. Another example was the harassment campaign targeting Alison Rapp that used #torrentialdownpour to organise itself. However, #torrentialdownpour has since been diluted by Twitter users applying it in an everyday context. Tweets linking the hashtag to targeting Rapp and/or Nintendo are still being made in 2020 alongside broader complaints about 'censorship,' including some from 2019 wishing the harassment campaign would come back.
13. They are also sites of conflict and strategic discourse: in 2020, 4chan responded to the sweeping Black Lives Matter protests by introducing competing hashtags like #WhiteLivesMatter and #WhiteOutWednesday. These were full of white-supremacist material and disinformation until

they were hijacked by fans of K-pop and flooded with K-pop videos (Andrews 2020; Molloy 2020). Those same fans then turned their attention to hashtags associated with the QAnon conspiracy, to the horror and outrage of QAnon members.

14. Roderick Graham has also written about how neo-Nazi and white-supremacist communities use hashtags as part of their online recruitment strategies—something that seems particularly relevant given the significant overlap between Gamergate and alt-right/neo-Nazi communities (Graham 2016).
15. Nominally, 'polite' enquiries become a form of harassment when combined with the inability for people online to 'get away' from their interrogators, particularly when they come in swarms—which the harassment community tries to ensure. This dynamic was highlighted by David Malki in the webcomic *Wondermark*, which lead to the process becoming known as 'sea-lioning' (Malki 2014; Massanari 2018, 2).
16. This strategy was more about leaving doubt in the minds of external observers and potential recruits as to who was at fault than preventing targets from knowing who was terrorising them. It was also never implemented with 100% effectiveness, since some of the community were very comfortable harassing under the #gamergate banner. The community's strategy for those cases was to claim that anyone doing so was a 'false flag' operation from external forces trying to frame them, again in an attempt to sow doubt and disorder.

References

Alexander, Julia. 2018. LGBTQ Creators Call out YouTube for Ongoing Homophobic Ads, Demonetization. *Polygon*, 4 June 2018. https://www.polygon.com/2018/6/4/17425686/lgbtq-creators-youtube-homophobic-ads-demonetization-pride.

Ambreen, Sam. 2019. FYI: Kiwi Farms Linked to At Least 2 Murders and 4 Suicides. *Left at the Lights* (blog). 8 March 2019. https://web.archive.org/web/20200312021829/https://samambreen.wordpress.com/2019/03/08/fyi-kiwi-farms-linked-to-at-least-2-murders-and-4-suicides/.

Andrews, Travis M. 2020. BTS Donates $1 Million to Black Lives Matter after K-Pop Fans Flood Hashtags to Support Movement. *Washington Post*, 7 June 2020. https://www.washingtonpost.com/technology/2020/06/07/bts-donation-k-pop-fans-black-lives-matter/.

'ArthurTheRef'. 2019. Turns Out Zak S. Is Worse Than We Thought. *Refereeing and Reflection* (blog). 11 February 2019. https://refereeingandreflection.wordpress.com/2019/02/11/turns-out-zak-s-is-worse-than-we-thought/.

Author Unknown. 2017. Comparison of Images Suggesting a Mistake Posting in Multiple Accounts. *Imgur*, 14 May 2017. https://imgur.com/a/qI5dJ.

Baio, Andy. 2014. 72 Hours of #Gamergate. *Medium*, 27 October 2014. https://medium.com/message/72-hours-of-gamergate-e00513f7cf5d#.p2o7p7y9m.

Bardo, Sal. 2018. YouTube Continues To Restrict LGBTQ Content. *Huffington Post* (blog). 16 January 2018. https://www.huffingtonpost.com/entry/youtube-continues-to-restrict-lgbtq-content_us_5a5e6628e4b03ed177016e90.

'@BechdelCast'. 2018. Don't Tell Anyone Someone Is Threatening You on the Internet or You'll Be Removed from the Internet: A Story in Three Actspic.Twitter.Com/6nlrYfaVI6. Tweet. *@BechdelCast* (blog). 8 October 2018. https://twitter.com/BechdelCast/status/1049477456296472581.

Beschizza, Rob. 2018. Twitter Apologizes for Saying Mailbomber's Threats Didn't Violate Its Rules. *Boing Boing*, 26 October 2018. https://boingboing.net/2018/10/26/rochelle-ritchie-complained-ab.html.

boyd, danah. 2011. Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In *A Networked Self: Identity, Community and Culture on Social Network Sites*, ed. Zizi Papacharissi, 39–58. New York; Abingdon: Routledge.

Bridle, James. 2017. Something Is Wrong on the Internet. *James Bridle* (blog). 6 November 2017. https://medium.com/@jamesbridle/something-is-wrong-on-the-internet-c39c471271d2.

———. 2018. *New Dark Age: Technology and the End of the Future*. London; New York, NY: Verso.

Buni, Catherine, and Soraya Chemaly. 2014. The Unsafety Net: How Social Media Turned Against Women. *The Atlantic*, 9 October 2014. https://www.theatlantic.com/technology/archive/2014/10/the-unsafety-net-how-social-media-turned-against-women/381261/.

Chess, Shira, and Adrienne Shaw. 2015. A Conspiracy of Fishes, or, How We Learned to Stop Worrying About #GamerGate and Embrace Hegemonic Masculinity. *Journal of Broadcasting & Electronic Media* 59 (1): 208–220. https://doi.org/10.1080/08838151.2014.999917.

———. 2016. We Are All Fishes Now: DiGRA, Feminism, and GamerGate. *Transactions of the Digital Games Research Association* 2 (2). http://todigra.org/index.php/todigra/article/view/39.

Citron, Danielle Keats. 2014. *Hate Crimes in Cyberspace*. Cambridge, MA: Harvard University Press.

'Coyote'. 2019. How Using Tumblr Is Undermining Your Community. *The Ace Theist* (blog). 23 November 2019. https://theacetheist.wordpress.com/2019/11/23/how-using-tumblr-is-undermining-your-community/.

D'Anastasio, Cecilia. 2018. How YouTube Fueled The Anti-Social Justice Movement. *Kotaku*, 20 September 2018. https://kotaku.com/how-youtube-fueled-the-anti-social-justice-movement-1829207455.

Ehrenkranz, Melanie. 2016. 4chan's New Racist Code: How Alt-Right Trolls Are Harassing Jews, Muslims and Black People. *Mic*, 3 October 2016. https://mic.com/articles/155739/4chan-new-racist-code-google-skype-skittle-alt-right-trolls-harass-jews-muslims-blacks#.fbZEctj4O.

Farokhmanesh, Megan. 2018. YouTube Is Still Restricting and Demonetizing LGBT Videos—And Adding Anti-LGBT Ads to Some. *The Verge*, 4 June 2018. https://www.theverge.com/2018/6/4/17424472/youtube-lgbt-demonetization-ads-algorithm.

Fleishman, Cooper, and Anthony Smith. 2016. (((Echoes))), Exposed: The Secret Symbol Neo-Nazis Use to Target Jews Online. *Mic*, 1 June 2016. https://mic.com/articles/144228/echoes-exposed-the-secret-symbol-neo-nazis-use-to-target-jews-online#.5hG0TbMEt.

Fogel, Stefanie. 2018. Video Game Developer Dies After Setting Herself on Fire. *Variety* (blog). 26 June 2018. https://variety.com/2018/gaming/news/chloe-sagal-death-1202858068/.

'@FreyjaErlings'. 2015. Thread of Images Documenting Zak Smith Encouraging Followers to Attack Targets. *Twitter*, 23 February 2015. https://twitter.com/freyjaerlings/status/569671505933705217.

Girdwood, Andrew. 2019. Blacklisted RPG Designer's Defamation Case Forces Ex-Girlfriend into Crowdfunding. *Geek Native*, 29 November 2019. https://www.geeknative.com/69638/backlisted-rpg-designers-defamation-case-forces-ex-girlfriend-into-crowdfunding/.

Golding, Dan, and Leena Van Deventer. 2016. *Game Changers: From Minecraft to Misogyny, the Fight for the Future of Videogames*. South Melbourne, VIC: Affirm Press.

Goodyear, Sheena. 2015. Meet The Woman Getting a PhD in Gamergate and the Death Eaters Trying To Stop Her. *The Mary Sue*, 15 June 2015. https://www.themarysue.com/phd-in-gamergate/.

Graham, Roderick. 2016. Inter-Ideological Mingling: White Extremist Ideology Entering the Mainstream on Twitter. *Sociological Spectrum* 36 (1): 24–36. https://doi.org/10.1080/02732173.2015.1075927.

Grey, Vivka. 2019. I'm Scared to Write This. *Facebook*, 13 February 2019. https://web.archive.org/save/https://www.facebook.com/VivkaCriesWolf/posts/2478145012257909.

Harper, Randi. 2016. Diving into the Cesspool: Fixing YouTube. *Medium*, 9 March 2016. https://medium.com/@randileeharper/diving-into-the-cesspool-fixing-youtube-3775a8afcd82#.7g9br3vfb.

Hatfield, Tom, and Cara Ellison. 2014. How Dungeons and Dragons Is Endorsing the Darkest Parts of the RPG Community. *Fail Forward*, 31 July 2014. https://web.archive.org/web/20200409210203/https://failforwardrpg.tumblr.com/post/93348768153/how-dungeons-and-dragons-is-endorsing-the-darkest.

Hern, Alex. 2017. YouTube Accused of 'Violence' against Young Children over Kids' Content. *The Guardian*, 7 November 2017, sec. Technology. https://www.theguardian.com/technology/2017/nov/07/youtube-accused-violence-against-young-children-kids-content-google-pre-school-abuse.

Hill, Olivia. 2017. White Wolf Hires Zak Smith—Im Out! *Onyx Path Forums*, 18 February 2017. http://forum.theonyxpath.com/forum/general/off-topic/1050412-white-wolf-hires-zak-smith-im-out/page4#post1051226.

Horwitz, Jeff, and Deepa Seetharaman. 2020. Facebook Executives Shut Down Efforts to Make the Site Less Divisive. *Wall Street Journal*, 26 May 2020, sec. Tech. http://archive.is/GPO6b.

Hudson, Laura. 2015. Curbing Online Abuse Isn't Impossible. Here's Where We Start. *Wired*, 15 May 2015. https://www.wired.com/2014/05/fighting-online-harassment/.

Hurley, Tracy. 2014a. What Happens When You Engage. *Sarah Darkmagic*, 1 August 2014. https://www.sarahdarkmagic.com/content/what-happens-when-you-engage.

———. 2014b. What Happens When You Engage—Continued. *Sarah Darkmagic*, 4 August 2014. https://www.sarahdarkmagic.com/content/what-happens-when-you-engage-continued.

———. 2014c. What Happens When You Engage—Act 3. *Sarah Darkmagic*, 5 August 2014. https://www.sarahdarkmagic.com/content/what-happens-when-you-engage-act-3.

Kim, Jeffrey, Elan Lee, Timothy Thomas, and Caroline Dombrowski. 2009. Storytelling in New Media: The Case of Alternate Reality Games, 2001–2009. *First Monday*. https://doi.org/10.5210/fm.v14i6.2484.

Kreider, Anna. 2014. Dangerous Hatred: Men Who Foment Misogyny in Geekdom. *Go Make Me a Sandwich*, 30 May 2014. https://gomakemeasandwich.wordpress.com/2014/05/30/dangerous-hatred-men-who-foment-misogyny-in-geekdom-twlong/.

———. 2015. This Post Is Insufferably Long, and I'm Sorry for That. *Go Make Me a Sandwich*, 10 March 2015. https://gomakemeasandwich.wordpress.com/2015/03/10/this-post-is-insufferably-long-and-im-sorry-for-that-longtw/.

Lewis, Rebecca. 2018. Alternative Influence: Broadcasting the Reactionary Right on Youtube. *Data & Society*. https://datasociety.net/output/alternative-influence/.

'lightninggrrl'. 2016. I Am Being Stalked and Harassed by Kiwi Farms and SA. *Wrong Planet*, 24 March 2016. http://wrongplanet.net/forums/viewtopic.php?t=308671.

Litt, Eden, and Eszter Hargittai. 2016. The Imagined Audience on Social Network Sites. *Social Media + Society* 2 (1). https://doi.org/10.1177/2056305116633482.

Lomas, Natasha. 2014. #Gamergate Shows Tech Needs Far Better Algorithms. *TechCrunch* (blog). 18 October 2014. http://social.techcrunch.com/2014/10/18/gamergate-tactics/.

———. 2017. Twitter's Abuse Problem Is Absolutely a Failure of Leadership. *TechCrunch* (blog). 12 October 2017. http://social.techcrunch.com/2017/10/12/twitters-abuse-problem-is-absolutely-a-failure-of-leadership/.

Lum, Zi-Ann. 2015. Veerender Jubbal, Sikh-Canadian Journalist, Wrongly ID'd As Paris Terror Suspect. *Huffington Post*, 16 November 2015. http://www.huffingtonpost.ca/2015/11/16/veerender-jubbal_n_8577520.html.

Malki, David. 2014. #1062; The Terrible Sea Lion. *Wondermark*, 19 September 2014. http://wondermark.com/1k62/.

Marwick, Alice E., and danah boyd. 2011. I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media & Society* 13 (1): 114–133. https://doi.org/10.1177/1461444810365313.

Masnick, Mike. 2017. Twitter Suspends Popehat For Writing About Violent Threats He Received from Another Twitter User. *Techdirt*, 3 August 2017. https://www.techdirt.com/articles/20170803/16341437919/twitter-suspends-popehat-writing-about-violent-threats-he-received-another-twitter-user.shtml.

Massanari, Adrienne L. 2015. #Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures. *New Media & Society*, October. https://doi.org/10.1177/1461444815608807.

———. 2018. Rethinking Research Ethics, Power, and the Risk of Visibility in the Era of the "Alt-Right" Gaze. *Social Media + Society* 4 (2). https://doi.org/10.1177/2056305118768302.

Matijevic, Paul. 2015. That Time Zak Smith Ran A Harassment Blog. *Ettin!* 2015. https://web.archive.org/web/20200409204001/https://www.tumbex.com/ettinjiggywithit.tumblr/post/106855388993/.

McGonigal, Jane. 2003a. *A Real Little Game: The Performance of Belief in Pervasive Play*. http://www.avantgame.com/MCGONIGAL%20A%20Real%20Little%20Game%20DiGRA%202003.pdf.

———. 2003b. *This Is Not a Game: Immersive Aesthetics and Collective Play*. http://www.seanstewart.org/beast/mcgonigal/notagame/paper.pdf.

Meifert-Menhard, Felicitas. 2013. *Playing the Text, Performing the Future: Future Narratives in Print and Digiture*. Berlin: De Gruyter. http://ezproxy.massey.ac.nz/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=661681&site=eds-live&scope=site.

Molloy, Parker. 2020. QAnon Followers Melt down after K-Pop Fans Take over Their Hashtag. *Media Matters*, 5 June 2020. https://www.mediamatters.org/qanon-conspiracy-theory/qanon-followers-melt-down-after-k-pop-fans-take-over-their-hashtag.

Nagel, Emily van der, and Jordan Frith. 2015. Anonymity, Pseudonymity, and the Agency of Online Identity: Examining the Social Practices of r/Gonewild. *First Monday* 20: 3. https://doi.org/10.5210/fm.v20i3.5615.

Nagy, Amanda. 2019. Please Feel Free to Share This Widely, on Any Platform You Have. *Facebook*, 11 February 2019. https://web.archive.org/web/20200410040514/https://www.facebook.com/story.php?story_fbid=10215845527064252&id=1027572040.

'Nerd City'. 2019. Youtube's Biggest Lie. *Nerd City*. https://www.youtube.com/watch?v=ll8zGaWhofU&feature=emb_logo.

Notopoulos, Katie. 2017. How Trolls Locked My Twitter Account for 10 Days, and Welp. *BuzzFeed News*, 2 December 2017. https://www.buzzfeednews.com/article/katienotopoulos/how-trolls-locked-my-twitter-account-for-10-days-and-welp.

O'Brien, Danny. 2009. Online Users Stick Claws into Torturer. *The Irish Times*. 20 February 2009. https://www.irishtimes.com/business/online-users-stick-claws-into-torturer-1.703999.

O'Donnell, Casey, and Mia Consalvo. 2015. Games Are Social/Media(Ted)/Technology Too…. *Social Media + Society* 1 (1). https://doi.org/10.1177/2056305115580337.

Olson, Dan. 2014. The Mods Are Always Asleep. *Medium*, 23 December 2014. https://medium.com/@FoldableHuman/the-mods-are-always-asleep-7f750f879fc#.tiq1cqq6d.

'pjamesstuart'. 2019. You Should Read This. *False Machine* (blog). 10 February 2019. https://falsemachine.blogspot.com/2019/02/you-should-read-this.html.

Pless, Margaret. 2015. 5 Reasons I Stand With Sarah Nyberg. *Internet Famous Angry Men*, 11 September 2015. http://idledillettante.com/2015/09/11/5-reasons-i-stand-with-sarah-nyberg/.

———. 2016. Kiwi Farms, the Web's Biggest Stalker Community. *New York Magazine*. 19 July 2016. http://nymag.com/selectall/2016/07/kiwi-farms-the-webs-biggest-community-of-stalkers.html.

Polansky, Lana. 2018. Worse than Scabs: Gamer Rage as Anti-Union Violence. *Rhizome*, 30 October 2018. http://rhizome.org/editorial/2018/oct/30/worse-than-scabs-gamer-rager-as-anti-worker-violence/.

RPG.net. 2013. Infraction for Zak (S: 17) Permanent Ban. *RPGnet Forums*, 5 August 2013. https://forum.rpg.net/index.php?threads/infraction-for-zak-s-17-permanent-ban.698051/.

'@ryanjreilly'. 2018. 'See u Soon Tick Tock'—Cesar Sayoc to Eric Holder a Few Weeks Ago. *Twitter* (blog). 26 October 2018. https://twitter.com/ryanjreilly/status/1055882833405886465.

Salter, Michael. 2017. From Geek Masculinity to Gamergate: The Technological Rationality of Online Abuse. *Crime Media Culture* 14 (2): 247–264.

Sarkeesian, Anita. 2015. One Week of Harassment on Twitter. *Feminist Frequency*, 27 January 2015. https://feministfrequency.com/2015/01/27/one-week-of-harassment-on-twitter/.

Sheldon, Lee. 2010. Ilovebees: Playing and Designing in Real-Time. In *Well Played 2.0: Video Games, Value and Meaning*, ed. Drew Davidson. ETC Press.

Smith, Zak. 2015a. Destroy. *Imgur*, 2015. https://i.imgur.com/NvBCbuf.png.

———. 2015b. Get At Him. *Imgur*, 2015. https://i.imgur.com/IKilzeI.png.

———. 2019. Official Announcements: I'm Suing. 12 June 2019. https://web.archive.org/save/https://officialzsannouncements.blogspot.com/2019/06/im-suing.html.

'Social Justice Viv'. 2016. Kiwi Farms, the Web's Biggest Community of... *Tumblr*, 13 September 2016. http://socialjusticeviv.tumblr.com/post/150364975017/kiwi-farms-the-webs-biggest-community-of.

Tabletop's Missing Stairs. 2019. Zak S and Other Horrible Tabletop People. *Tabletop's Missing Stairs* (blog). 11 February 2019. https://tabletopsmissing-stairs.blogspot.com/2019/02/zak-s-and-other-horrible-tabletop-people.html.

'Tenkar'. 2019. Zak S Sues Mandy in Canadian Court for Defamation—Probably Because Its Much Easier for the Plaintiff to Get a Favorable Ruling (Not Enforceable in the US). 14 June 2019. https://www.tenkarstavern.com/2019/06/zak-s-sues-mandy-in-canadian-court-for.html.

Trepte, Sabine. 2015. Social Media, Privacy, and Self-Disclosure: The Turbulence Caused by Social Media's Affordances. *Social Media + Society* 1 (1). https://doi.org/10.1177/2056305115578681.

'violetblue'. 2018. Why PayPal's Crackdown on ASMR Creators Should Worry You. *Engadget*, 14 September 2018. https://www.engadget.com/2018/09/14/paypal-ban-asmr-sound-art-therapy/.

West, Lindy. 2014. Twitter Doesn't Think These Rape and Death Threats Are Harassment. *Daily Dot*, 23 December 2014. http://www.dailydot.com/opinion/twitter-harassment-rape-death-threat-report/.

Williams, Jane. n.d. Do Reblogs Disappear If a Tumblr Is Deleted? *It Still Works*. Accessed 5 November 2018. https://itstillworks.com/12760452/do-reblogs-disappear-if-a-tumblr-is-deleted.

Wilson, Jason. 2016. Game Developer Group Doesn't Let Nintendo off the Hook for Firing Alison Rapp. *Venture Beat*, 5 April 2016. http://venturebeat.com/2016/04/05/game-developer-group-doesnt-let-nintendo-off-the-hook-for-firing-alison-rapp/.

CHAPTER 5

Problematic Tools and Platform Complicity

Where the focus of the last two chapters was exploring ways that the dynamics and behaviour of harassment communities match those of alternate reality games (ARGs), the next two chapters shift their attention to possible tools for limiting the impact of harassment in online spaces. Firstly, this chapter will discuss tools currently being used to fight harassment, alongside ones commonly proposed as promising future solutions, and explore ways that they can introduce as many problems as they solve. It will then argue that these tools, both current and proposed, are popular because of the underlying philosophies and business models of social media companies, which are complicit in—and sometimes directly profit from—the abuse unfolding across and through their platforms.

Problematic Tools

As discussed in Chap. 1, the philosophical foundations of the internet are grounded in a technolibertarian ethos that sees humans as atomised individuals rendered equally disembodied and 'equal' in digital spaces (Massanari 2015, 5; Shepherd et al. 2015, 4; Turner 2006). Rubin et al. argue that this framework allows social media platforms and technology companies to hide behind a 'façade of neutrality' that ignores disparate impact on marginalised groups (Rubin et al. 2020, 1). This cultural logic surfaces in many ways within proposals to deal with online harassment

K. Veale, *Gaming the Dynamics of Online Harassment*,
https://doi.org/10.1007/978-3-030-60410-3_5

which are either non-functional, dysfunctional or actively harmful to marginalised groups online.

One such example is that a rich strain of 'technological solutionism' (Morozov 2013) exists within the owners and designers of social network platforms. As a result, R. Stuart Geiger critiques both them and their work as treating social and cultural problems as technological problems that can be solved with equally technological solutions[1] (Geiger 2016, 791). This is also a logic that justifies vague attempts to 'develop' a solution to online harassment while doing nothing concrete in the meanwhile, because 'the tools are not available yet.'

Another motivator which overlaps with technological solutionism is the reality of late-stage capitalism. Both a disinterested lack of action and low-energy attempts to seek automated solutions are partly motivated by a simple desire to raise profits by saving money, potentially by not spending it at all, and instead convincing people that online harassment is an unavoidable inevitability. However, profit-seeking capitalist logics are corrosive to any approach to online harassment and abuse[2] because it puts humans last—in terms of both what 'acceptable' results are achieved and how they are to be achieved. For example, Facebook is unique in employing human personnel to personally review reports of abuse and harassment, which is a significant improvement from attempting to solve the problem using algorithms. However, those moderators are predominately hired at very low wages on the international market and then provided no support when they develop Post-Traumatic Stress Disorder (PTSD) from their exposure to horrific content that leaves them at high risk of suicide (Dwoskin 2018; Whigham 2018). The capitalist context of internet platforms means what could be one of the best solutions to online moderation treats its staff as disposable, even as the platform makes record profits (Constine 2019).[3]

Leaving aside how capitalism would attempt to justify the concept, all attempts to solve crowdsourced terrorism and online harassment algorithmically are fundamentally doomed to failure, as discussed in Chap. 4. Harassment communities learn to game the algorithms and will use them as tools for further terrorising marginalised groups. Nicholas Diakopoulos advocates for creating frameworks for assessing 'algorithmic accountability,' in order to 'articulate the power structures, biases and influences that computational artifacts play in society' (Diakopoulos 2015, 399). In addition to algorithmic accountability, I argue that there is an equal need for *affordance accountability*, where online spaces such as social networks can be assessed for how their affordances impact the networked publics that

they mediate. A sentiment that several people who wish to remain anonymous have expressed is that in general, technology industries need to ask how their tools might be used by the worst-possible people, rather than assuming they will be blandly neutral tools. The assumption of technological neutrality grows from the same technolibertarian roots as technological solutionism: technology affects 'everyone equally' because the people who are the predominate designers of online tools imagine their audience to be the presumed cultural default of straight, cisgender, white men:

> Platform policies and tools are largely designed for a presumptively homogenous pool of users, without recognition of the differing impacts and experiences of specific individuals or historically marginalized groups. This façade of neutrality has made it especially difficult for major social media platforms to manage surges of harassment fueled misogyny or racism. (Rubin et al. 2020, 1)

This produces dynamics where technology and online tools are imagined for how they might impact the safest demographics, and any problems faced by people outside of the presumed cultural default are aberrant and do not motivate substantial change.[4] The assumption that technology is neutral results in internet-enabled devices becoming tools for twenty-first-century domestic abuse or burglary (Bowles 2018; Denne et al. 2018; Vella 2018), and insurance companies reframing themselves around only serving customers with health-trackers (Barlyn 2018; Beschizza 2018).

Algorithmic solutions to harassment might be cheaper for social media platforms and other online spaces, but they fundamentally do not work and only offer new tools to be subverted and used for crowdsourced terrorism. Instead, approaches that proactively engage with both algorithmic and affordance accountability are necessary, and that will involve some of the strong, assertive leadership and diversity-building that Golding and Van Deventer note is currently lacking from technology companies (Golding and Van Deventer 2016, 101).

Blockbots

Bot-based collective blocklists, aka blockbots, are another tool that has been subverted and turned against marginalised groups—and much as with algorithms, the problem is the social engagement with the

technology rather than the technology in a vacuum. Blockbots allow users to subscribe to lists curated by individuals, communities and/or algorithms to automate the otherwise futilely time-consuming process of blocking the members of a harassment campaign and deal with the disparities of scale involved (Geiger 2016, 788–89, 795). R. Stuart Geiger has written extensively about the impact of blockbot technology on Twitter. Geiger stresses that the developers and the community of users had ongoing discussions about how to mitigate their shared concerns about the negative impacts of people being added to a blocklist by accident or for inappropriate reasons (Geiger 2016, 797, 799). Unfortunately, although Geiger argues clearly that blockbots 'should not be seen as the kind of top-down technical solution that can be installed to fix the problem "once and for all,"' (Geiger 2016, 799), that is exactly how some of the online community have treated them. The initial genesis of the problem began innocently when Randi Harper popularised the 'ggautoblocker' as a safe response to Gamergate: a blockbot which has been noted as significantly improving the lives of many and which had over 3000 subscribers by mid-2015 (Geiger 2016, 795–96). Although the ggautoblocker itself was carefully curated and transparent, and had a formal system for handling appeals involving multiple moderators, Randi Harper also circulated her own 'personal' blocklist over social media, claiming that it was full of 'bad actors' in addition to the normal ggautoblocker list.

Unfortunately, the online community approached this new list *exactly* as a top-down solution that could be trusted without scrutiny, despite the fact it lacked all of the support/moderation/appeal structures that Geiger argues are necessary (Geiger 2016, 797–98). If someone who seems reliable presents a list of many thousands of 'bad people,' described as harassers, neo-Nazis, Men's Rights Activists and other serious online problems, very few people are motivated to check ('@persenche' 2020; 'sjwomble' 2016). Instead, it is far easier to take the legitimacy of the list in good faith and subscribe to it.

In fact, since the list was her own personal one, it was curated by Harper's whim. It included a significant number of trans people who disagreed with her on trans issues—and the people who spoke up on their behalf (Jhaver et al. 2018, 24; 'secretgamergirl' 2018; 'sjwomble' 2016). Instantly, members of a significantly marginalised group already targeted by harassment campaigns and who disproportionately depend on online environments (Jhaver et al. 2018, 18) were cut off from parts of the online community, and unaware of why they were suddenly invisible. The

people subscribing to Harper's personal blocklist were unaware of the impact it would have. If someone you followed went 'radio silent' after you subscribed to her blocklist, it would seem like they had just chosen to stop posting. There was nothing to signal that the blocklist had caused the change. The problem magnified dramatically when celebrities such as Wil Wheaton incorporated Harper's personal blocklist into their own personal lists, before advertising those lists to their broad audiences as an 'improvement' on the ggautoblocker. This resulted in a process of 'metastasization' whereby uncurated blocklists that functioned as infrastructural silencing tools propagate through personal and professional networks (Jhaver et al. 2018, 23; '@persenche' 2020; 'secretgamergirl' 2018; 'sjwomble' 2016).

> It's not the blocklist, it's the way they metastasize. You sub to a blocklist, and then that becomes the basis for your own list, and people who were randomly swept up keep getting carried forward with each new iteration of the list.
>
> And once you use a blocklist? *you can never undo it*.
>
> and you have *no idea* who you blocked.
>
> I've had actual IRL friends discover they had me blocked.
>
> You are surrendering control over the scope of your interactions to a *stranger* and whatever biases and personal grudges they happen to embed in their list. ('@persenche' 2020)

Worse, once the impact of the problem became known, harassment communities—particularly anti-trans bigots—began adding people from the trans community and other marginalised groups to lists that lacked moderation and scrutiny. As a result, many people have found themselves silenced without knowing why, or even necessarily, being able to prove that is what has happened to them ('@persenche' 2020).

There have been attempts to use blockbot technology to solve the problem. For example, @unblock_list will inform users what blocklists they are on if they tweet 'what blocklists' at the bot, and has lists of people it recommends as unfairly included on blocklists. However, that does not remove affected people from the lists, and if you as a user decide that you have subscribed to a list that you now believe to be corrupt, there is no way to simply unsubscribe without removing people manually. There is no method to solve the disproportionate visibility and ongoing contagion provided by Harper's blocklist. It is still being circulated and added to in

2020, without any oversight since it is her personal list, and the online community has no reason to distrust its contents. Plus, it is still propagating through other people incorporating it into their own blocklists and sharing them. Additionally, legitimate bad actors are also included on the unblock_list among the people unfairly targeted, and there is no way to tell who is who without checking them individually or knowing them beforehand. Effectively, once blocklist/blockbot technology has been abused, it is difficult to restore trust back into the system and impossible to generate a 'clean' list.

Blockbot technology presents a strong enough tool *in theory* that there is interest in learning how to rehabilitate them and rendering them safe for marginalised groups to enjoy their benefits without being targeted through them. For example, Jhaver et al. present a substantial analysis of blockbot dynamics and argue that there are ways which could mitigate problems like the ones discussed here going forward (Jhaver et al. 2018). These suggestions exist alongside Geiger's suggestions for responsible curation (Geiger 2016, 797–98). However, users from marginalised groups have argued that there is a fundamental and foundational mistrust in the technology, based on how easily and insidiously it has been used against them ('@persenche' 2020; 'sjwomble' 2016). It may be that it is as difficult to restore trust into blockbot technology as it is to salvage a compromised blocklist.

'Real Name' Policies

Another suggestion often made for mitigating problems with harassment online is 'real name' policies and other initiatives that assume that harassment only happens due to the anonymity of the environment. The core concept that removing anonymity will reduce online harassment has undergone significant critique: Emily van der Nagel and Jordan Frith argue that Facebook's strong opposition to anonymity is motivated by profit rather than security (van der Nagel and Frith 2015); some argue that these policies will have (and are already having) a disproportionate impact on already marginalised groups (Carroll and Holpuch 2015; Cat 2015; Citron 2014, 239; Holpuch 2015); while others argue that the approach fundamentally will not work because people already harass under their legal names, and/or the policies misunderstand the motivations for online harassment (Collins and Cross 2015; Cross 2015; Dash 2016; Phillips 2015, 156; Rösner and Krämer 2016). Removing anonymity as a

tool would likely have a bigger negative impact on the people harassment communities attack than on the people doing the attacking. Whitney Phillips argues that focusing on behaviour is a more productive option than attempting to remove anonymity overall:

> I believe we should focus on specific problematic behavior and the context in which the behaviors occurred. In each circumstance we should examine who is exerting power over whom, what the repercussions are for the target and whether the behaviors are persistent or ephemeral. By focusing on specifics, one is much more likely to arrive at nuanced conclusions, and more important, actionable steps. (Phillips 2014)

The Free Network

In another example of tools that can potentially be dangerous if misunderstood, The Free Network (TFN) represents a suite of interconnected digital platforms that can create networked publics and are often suggested as a safer alternative form of social media. Robert W. Gehl has written extensively about the significant differences between the networked publics associated with alternative social media (ASM) when compared to corporate social media (CSM), and the ways that ASM are more responsive to the needs of their audiences (Gehl 2015). However, the underlying design of platforms within TFN such as Mastodon, Diaspora and others[5] (Tilley 2017) opens the door to security and harassment problems. These problems are caused by how their infrastructure functions and how the public conceptualises both social networks and networked publics ('@3fingeredfox' 2018; '@adrienneleigh' 2017a; 2017b; 2018; '@cassolotl' 2018; Pincus 2017; '@_sagesharp_' 2018). In essence, even though a platform like Reddit hosts innumerable networked publics,[6] its underlying infrastructure is consistent across all of them. When dealing with an example of ASM that uses TFN, such as Mastodon, that is no longer true.

TFN is a platform for creating networked publics that can communicate with each other, rather than a singular social network. The affordances and design of one space created using Mastodon may be entirely distinct from any others, contributing to differences in the networked publics and cultures of their environments. This dynamic produces many strengths: there is a significant diversity of options to choose between compared to more monolithic CSMs such as Twitter. In addition, people are free to both create and find spaces that will suit their needs, all of

which are not driven by a profit motive and advertising. The downsides are that

1. the people running each networked public have total control over them;
2. there are no tools for oversight or transparency of those in charge; and
3. the technology itself provides the people running each networked public with access to any communication internal to their server or entering it—private or not (Anthony 2018).

The underlying code of each instance might share the same software backbone, but can be modified by the administrators at will. For example, they could set up their instance to ignore commands to delete posts, keeping them indefinitely—and again, regardless of whether they were private or not. As Adrienne Leigh says,

> to folks helpfully pointing out that Twitter has copies of all your Twitter PMs, etc.—yes, users understand that, i promise.
>
> Here's the thing: Twitter is untrustworthy but THERE'S ONLY ONE OF IT. And if it changes management, WE'LL KNOW. GNUsocial instances are myriad, ANYONE can run them, & they can be handed off silently to new admins. Instead of trusting ONE shady entity you are trusting MANY. ('@adrienneleigh' 2017a)

One of the important things that Leigh highlights here is the issue of perception. Many of the issues with TFN could largely be resolved if the users were clearly aware that it is not a singular monolithic networked public and instead provides tools that can produce networked publics run by random people with their own rules. There are different clear risks—and approaches to manage those risks—when being social within a server run by someone in your circle of friends or a stranger online when compared to CSM. The issue is that currently it is not clear that TFN produces spaces that need to be treated as closer to the former than CSM, meaning people do not manage the risks involved appropriately.

Many people are unaware of how the platforms connected to TFN function. It is easy to conclude that they are subscribing to a singular networked public called 'Mastodon' rather than a singular splinter run by volunteers, without oversight, which communicates to the wider collective ('@adrienneleigh' 2017a). 'Counter.Social' provides an example of the problem. The site looks professional and presents itself as a 'social network

platform,' suggesting a similar level of organisational structure to a CSM ('@th3j35t3r' 2017). However, the site is functionally a Mastodon instance run and entirely controlled by a hacker known as 'The Jester' (@th3j35t3r) who has previously gained access to personal information in order to hunt and attack specific targets (Ullrich 2012; Wagenseil 2012). The site's design is solely at The Jester's whim and blocks accounts from a list of 'hostile nations' ('@th3j35t3r' 2017) on the grounds that they are host to threats to online privacy and security. It additionally blocks Pakistan because The Jester blames the citizens of the country for the fact Osama Bin Laden hid there ('@th3j35t3r' 2018). There is nothing in the site's design to inform the public that it exists under the control of one person with no oversight because it looks like CSM.

An issue that magnifies the problem is that Mastodon and Diaspora, two of the larger sub-platforms within TFN, lack basic tools for handling harassment because just as with CSM, the developers do not take the issue seriously ('@3fingeredfox,' 2018; '@adrienneleigh,' 2018).

The fundamental design of TFN and its multiple subsidiaries is also open to subversion in other ways. A pragmatic consequence of every instance built using the different TFN platforms functioning as an island unto itself is that it is trivially easy to take someone's username and identifying information from one instance and register an account on a different one ('kovah' 2017). The problem here is, again, a social one. People intuitively understand by now that a particular username on a Yahoo email account is not necessarily the same person who is using that username for their Gmail account. However, since CSMs like Twitter are a singular networked public, people often do not apply the same logic they use for email to ASMs like Mastodon instances, Diaspora pods and other equivalents.[7] As a result, they are ripe spaces for impersonating someone through registering their name on another instance, and attempts to destroy someone's reputation in this way have already happened.[8] Likewise, someone could use exactly the same process to attack people, but deny responsibility by claiming they had been impersonated. From the outside, the two cases would look identical, and despite using all of their identifying information, it would be functionally impossible to prove in either direction without external information.

Platform Complicity

A fundamental elephant in the room regarding online harassment is that social network platforms are complicit in the abuse done using their services. They provide the tools and spaces with which abuse and harassment are committed, and this fact must be clearer to them than anyone without insider access to their traffic and other information. However, they are—at minimum—indifferent to the problem. Both CSMs and many examples of ASMs are designed and dominated by the demographics of straight, cisgender white men who are at the least significant statistical risk of online harassment (Rubin et al. 2020, 1). As a result, there is a lack of consideration for how the affordances of online spaces will be weaponised, and a corresponding lack of motivation to become involved once they *have* been weaponised. There are parallels to what Danielle Keats Citron identifies regarding law enforcement and the legal system's disinterest in prosecuting abuse and harassment (Citron 2014, 73–94). Male-dominated organisations do not establish rules to account for concerns that do not affect them, and are then unmotivated to deal with problems that either breach the rules which exist or suggest those rules need to be changed. However, the complicity of CSMs and the people behind them goes beyond disinterest.

There are tools available to social networks that are not being applied or at least not applied to the service of reducing abuse. A very simple example is that despite claims Twitter itself does not have the tools to combat abuse and harassment within its network; Twitter has rapidly removed gifs, images and footage of the Olympic Games because of copyright claims from the Olympic Committee (Silverman 2016). This shows that Twitter has access to tools that would allow it to take an active hand in shaping the networked public that it is responsible for, but refuses to do so outside of protecting financial relationships with corporate partnerships.[9] If Twitter is capable of pulling down Olympic gifs, it is capable of pulling down gifs designed to trigger epilepsy and cause seizures which were used to attack Jewish journalist Kurt Eichenwald, among many others (Eichenwald 2016). Likewise, the same example suggests Twitter has the tools to remove images from the Holocaust that people protesting neo-Nazis are flooded with (Warzel 2016b). The rules and guidelines by which Twitter operates, and even the terms and services, are also frequently not applied, even in clear-cut cases (Sarkeesian 2015; West 2014; Warzel 2016a). Another example is the fact that Twitter is legally required

to block neo-Nazi accounts within France and Germany and created a tool to do so in 2012, but refuses to apply the same tool to its networked public globally (Feiner 2019; Lomas 2017; Martin 2017). We are left to speculate as to why a tool that would be easier in some ways to apply globally has instead been applied to the lowest number of countries required by law.[10] As discussed earlier, Facebook has human employees review reports of abuse and harassment, which is a significant improvement from attempting to solve the problem using algorithms. However, Facebook publishes rules for moderation that requires those employees to adhere to rules which dictate that actively hateful, clearly anti-Semitic imagery is acceptable and must not be interfered with (Koebler and Cox 2018). These rules are part of a historical problem where Facebook monetises extremist white-supremacist content and creates rules by which it can continue to do so:

> In 2018, Channel 4's Dispatches uncovered Facebook's 'shielded review' moderation system which permitted high-traffic right wing hate speech posted by groups such as the English Defence League and Britain First to remain online despite numerous complaints. (Thompson 2019, 84)

Critics have argued that one of the reasons toxic content is permitted on the platform is because the people posting it and engaging with it are 'the really valuable ones' to the site's bottom line (Schipp 2018).[11]

Although there is not a proven explanation for the failure of CSMs to apply tools they have available to limit harassment and crowdsourced terrorism, there are concrete examples where social media companies have directly profited from harassment on their networks (Shepherd et al. 2015, 5). Adrienne Massanari highlights that during an event called 'The Fappening,' where a collection of private celebrity photos were posted online, people on Reddit directly sponsored the circulation of these images, buying enough of Reddit's internal currency in six days to run the site for a month (Massanari 2015, 8). On Twitter, 'impressions' are defined as tweets that 'actually generate interaction or replies from others online' (Doctor 2013). Given that the people targeted by harassment have high levels of impressions as a direct result of their harassment, they and the people harassing them are valuable to the network so long as the harassment is sustained (Monteiro 2018; Vance 2014). A 'quote tweet' is a kind of retweet that makes the content you are responding to visible to your followers. This is frequently used to highlight material in order to

critique it or highlight its problematic nature: however, Twitter's algorithmic systems seem to reward that kind of engagement as if a quote tweet was warm approval. As a result, the same systems which remove context from what 'engagement' is produces dynamics where people promote the content they set out to critique, because of the business model underlying those systems.

Facebook's algorithmic business models are equally implicated in these problems:

> Under Facebook's engagement-based metrics, a user who likes, shares or comments on 1500 pieces of content has more influence on the platform and its algorithms than one who interacts with just 15 posts, allowing "super-sharers" to drown out less-active users. (Horwitz and Seetharaman 2020)

As discussed in Chap. 4, executives within the company[12] scuttled attempts to remedy these problems despite research internal to the platform highlighting that

- accounts with hyperactive engagement were far more partisan on average than normal Facebook users;
- they were more likely to behave suspiciously, sometimes appearing on the platform as much as 20 hours a day and engaging in spam-like behaviour;
- partly as a result of algorithms influenced by these accounts, 64% of the growth in extremist groups occurred *because Facebook's algorithms recommended them* (Horwitz and Seetharaman 2020).

YouTube's systems[13] extend this dynamic further by opening the door to monetised abuse, on top of the algorithmic systems that encourage and profit from it that have been discussed previously: YouTube connects paid anti-LGBTQA+ advertising to LGBTQA+ content, encouraging bigots to pay in order to harass people producing content relevant to rainbow communities online (Alexander 2018; Bardo 2018; Farokhmanesh 2018). In addition, Rebecca Lewis discusses the example of YouTube's Super Chat, where users can pay to have their comments highlighted and pinned in a comment stream to increase their visibility—often including extremist or harassing content in the process (Lewis 2018, 41–42).

YouTube's relationship with extremism and harassment, however, goes beyond simply monetising what is already there: YouTube's networked public actively encourages and rewards the creation and distribution of radicalising extremist material. Zeynep Tufekci documented the extent to which YouTube's recommendation algorithms ensure audiences encounter more and more extreme content over time in an attempt to keep them watching, regardless of what that content is, and drive profit.

> I experimented with nonpolitical topics. The same basic pattern emerged. Videos about vegetarianism led to videos about veganism. Videos about jogging led to videos about running ultramarathons.
>
> It seems as if you are never "hard core" enough for YouTube's recommendation algorithm. It promotes, recommends and disseminates videos in a manner that appears to constantly up the stakes. Given its billion or so users, YouTube may be one of the most powerful radicalizing instruments of the 21st century.
>
> This is not because a cabal of YouTube engineers is plotting to drive the world off a cliff. A more likely explanation has to do with the nexus of artificial intelligence and Google's business model. (YouTube is owned by Google.) For all its lofty rhetoric, Google is an advertising broker, selling our attention to companies that will pay for it. The longer people stay on YouTube, the more money Google makes. (Tufekci 2018)

As a result, audiences who encounter material from what Rebecca Lewis calls an 'alt-right influencer network' (AIN) are gradually funnelled into more and more extreme content and discourse (Lewis 2018). Lewis' analysis highlights that the white-supremacists and neo-Nazis who prefer the more neutral branding of 'alt-right' are adept at search engine optimisation, making their work very visible to searches that connect to their subject areas, at the same time as colonising more neutral or progressive language in order to expand their audiences.

> I also found that influencers are explicitly using terminology affiliated with progressive social justice movements and are therefore appearing in search results for those terms. A number of popular videos from conservative influencers use the terms "social justice," "liberal," and "leftist" in their video titles, as well as more specific terms like "intersectionality." Currently, a YouTube search for any of those terms will bring back content from conservative and libertarian political influencers within the top 10 YouTube search results. (Lewis 2018, 31)

However, Lewis is very clear that although the algorithmic dimensions of YouTube's problematic networked public are important, they are not the only problem. She argues that the neo-Nazi influencer network has developed social strategies designed to share their audiences and boost the viewership of all of those participating, even without algorithmic support (Lewis 2018, 36–39). As a result, a solely technological solution will not be enough to sanitise the networked public (Lewis 2018, 36). Another problem without a solely technological solution is the fact that YouTube's networked public is not a one-way system, where the influencers are isolated from the perspectives of their audiences:

> The easy feedback systems on YouTube lead to discursive loops, in which influencers build audiences that ask for, or reward, certain types of content. For many of the political influencers in the AIN, the more extremist content they make, the more of an extremist and dedicated audience they build. Such audiences can, in turn, drive political influencers to deliver ever more extreme content. (Lewis 2018, 40)

Just as with any networked public, the design of YouTube's algorithms is a substantial force-multiplier to the dynamics evolving within the space. However, removing the algorithms in isolation would do nothing to deal with the fact that there are online extremists purposefully radicalising their audiences through a variety of social practices, and that audience dynamics encourage influencers to make more extreme content as well.

This chapter has explored the ways that tools for responding to online harassment are designed and imagined by people who are among the least likely to experience harassment themselves. Those same people are among the most likely to find the idea of trying to solve complex social issues with 'neutral' technology appealing. As a result, proposals to deal with online harassment are often either non-functional, dysfunctional or actively harmful to marginalised groups online.

In addition, technology industries and social network platforms have financial motivations not to change the status quo. We can see signs of this through occasions where corporate social media have

- not applied existing tools to the service of reducing harassment on their platforms;
- profited directly from harassment and not changed the infrastructure that enabled that dynamic to happen;

- been resistant to stopping harassment as a result of those profits or
- designed new services to maximise that profit.

The question becomes, if the people behind networked publics are either indifferent to—or profiting from—harassment and crowdsourced terrorism, what can be done to prevent or limit it and its impacts in future? The next chapter will propose methods for disrupting harassment communities and blunting their impact through treating them as ARGs.

Notes

1. Perhaps unsurprisingly, those who embrace 'technological solutionism' have a demonstrated tendency not to engage with people working in the humanities and social sciences on those exact problems, in all their depth and nuance.
2. Alongside being corrosive to functionally any human endeavour whatsoever.
3. In 2020, the live-streaming service Twitch established a Safety Advisory Council (SAC) to discuss moderating the community and drew public attention to its membership in ways that exposed the people involved to a harassment campaign (Grayson 2020). Having placed those people at risk, Twitch was slow to respond to the problem, took only lacklustre steps to protect the people on the SAC when they did act, and spent more time emphasising that members of the SAC were not Twitch staff and could not make binding decisions. Overall, the clear emphasis was more focused on protecting Twitch as a brand by mollifying the harassment campaign they had given targets than defending the people they had made vulnerable.
4. These issues affect culture more broadly, as can be seen in Citron's examination of how extensively both the police and the legal system fail to take online harassment and threats seriously or to use tools that are already available for handling those cases (Citron 2014, 73–91).
5. Both saw a surge in accounts after the closure of Google Plus in 2019.
6. Subreddits develop their own cultures, shaped by and responding to the design and affordances of Reddit's infrastructure.
7. Particularly not when examples like Counter.Social are designed to encourage people to mistake them for something closer to CSM.
8. Another example where I have witnessed the behaviour personally, but it has not left documentable material behind.
9. A report from the UK government highlighted that Google has similar issues: 'We note that Google can act quickly to remove videos from YouTube when they are found to infringe copyright rules, but that the same prompt action is not taken when the material involves hateful or illegal content' (House of Commons/Home Affairs Committee 2017, 10, 21).

10. Similarly, Twitter has had great success in removing content from ISIS supporters across its network, but has refused to apply the same tool to white-supremacist extremism on the grounds that doing so would filter Republicans in the United States (Cox and Koebler 2019). In 2020, a Twitter account resharing content from Donald Trump with no alterations was suspended after operating for 68 hours on the grounds it was 'glorifying violence,' while Trump's account itself was defended as 'public interest' (Yeo 2020).
11. And as discussed earlier, Facebook's argument that anonymity drives harassment is motivated by profit rather than security (van der Nagel and Frith 2015).
12. Leaks have revealed the concerns of rank-and-file Facebook employees about the role of the platform in boosting dangerous content from Donald Trump which is in breach of Facebook's guidelines but which is allowed to circulate because of his prominence (Newton 2020). They communicated these concerns to Facebook executives, who did not respond.
13. This discussion is leaving aside the ways YouTube's algorithms effectively boost and automate harassment, as discussed in Chap. 4.

References

'@3fingeredfox'. 2018. Critiques Structure and Abuse Tools on Mastodon. Tweet. *@3fingeredfox* (blog). 16 August 2018. https://twitter.com/3fingeredfox/status/1030218115848781824.

'@adrienneleigh'. 2017a. So, Mastodon. Just FYI, If a SINGLE Hostile User on a SINGLE Malicious Instance Follows You, All Your Posts May Be Kept Forever—. Tweet. *@adrienneleigh* (blog). 6 April 2017. https://twitter.com/adrienneleigh/status/850061121184604160.

———'. 2017b. Critiques of Mastodon's Governance and Moderation Structures. Tweet. *@adrienneleigh* (blog). 8 November 2017. https://twitter.com/adrienneleigh/status/928422479361490944.

———'. 2018. THREAD: In Light of so Many Folks Moving to Mastodon/the Fediverse, I Want to Re-up a Couple of Threads I Did Last Year about Safety and Other Issues. None of These Have Gone Away, and the Increased Volume Is Gonna Make Them Worse, Not Better. Tweet. *@i* (blog). 16 August 2018. https://twitter.com/i/web/status/1030213888850030592.

Alexander, Julia. 2018. LGBTQ Creators Call out YouTube for Ongoing Homophobic Ads, Demonetization. *Polygon*, 4 June 2018. https://www.polygon.com/2018/6/4/17425686/lgbtq-creators-youtube-homophobic-ads-demonetization-pride.

Anthony, Noëlle. 2018. An Increasingly Less-Brief Guide to Mastodon. *GitHub*, 3 October 2018. https://github.com/joyeusenoelle/GuideToMastodon.

Bardo, Sal. 2018. YouTube Continues To Restrict LGBTQ Content. *Huffington Post (blog)* (16 January 2018) https://www.huffingtonpost.com/entry/youtube-continues-to-restrict-lgbtq-content_us_5a5e6628e4b03ed177016e90.

Barlyn, Suzanne. 2018. Strap on the Fitbit: John Hancock to Sell Only Interactive Life... *Reuters*, 19 September 2018. https://www.reuters.com/article/us-manulife-financi-john-hancock-lifeins/strap-on-the-fitbit-john-hancock-to-sell-only-interactive-life-insurance-idUSKCN1LZ1WL.

Beschizza, Rob. 2018. Major U.S. Insurance Company to Sell Only Health-Tracker Backed Life Insurance. *Boing Boing*, 20 September 2018. https://boingboing.net/2018/09/20/major-insurance-company-to-onl.html.

Bowles, Nellie. 2018. Thermostats, Locks and Lights: Digital Tools of Domestic Abuse. *New York Times*, 23 June 2018, sec. Technology. https://www.nytimes.com/2018/06/23/technology/smart-home-devices-domestic-abuse.html.

Carroll, Rory, and Amanda Holpuch. 2015. Hold the Applause for Facebook's Rainbow-Colored Profiles, Activists Say. *The Guardian*, 28 June 2015. http://www.theguardian.com/world/2015/jun/28/facebook-rainbow-colored-profiles-san-francisco-pride.

'@cassolotl'. 2018. 'I Left Mastodon Yesterday'. *Cassian* (blog). 4 June 2018. https://medium.com/@cassolotl/i-left-mastodon-yesterday-4c5796b0f548.

Cat, Zoë. 2015. My Name Is Only Real Enough to Work at Facebook, Not to Use on the Site. *Medium* (blog). 27 June 2015. https://medium.com/@zip/my-name-is-only-real-enough-to-work-at-facebook-not-to-use-on-the-site-c37daf3f4b03.

Citron, Danielle Keats. 2014. *Hate Crimes in Cyberspace*. Cambridge, MA: Harvard University Press.

Collins, Katie, and Katherine Cross. 2015. Why Outlawing Anonymity Will Not Halt Online Abuse. *Wired*, 19 August 2015. http://www.wired.co.uk/article/real-name-policies-anonymity-online-harassment.

Constine, Josh. 2019. Facebook Shares Rise on Strong Q3, Users up 2% to 2.45B. *TechCrunch* (blog). 30 October 2019. https://social.techcrunch.com/2019/10/30/facebook-earnings-q3-2019/.

Cox, Joseph, and Jason Koebler. 2019. Twitter Won't Treat White Supremacy Like ISIS Because It'd Have to Ban Some GOP Politicians Too. *Vice* (blog). 25 April 2019. https://www.vice.com/en_us/article/a3xgq5/why-wont-twitter-treat-white-supremacy-like-isis-because-it-would-mean-banning-some-republican-politicians-too.

Cross, Katherine. 2015. Causes of Online Harassment. In Malmö, Sweden. http://videos.theconference.se/katherine-cross-causes-of-online-harassment.

Dash, Anil. 2016. The Immortal Myths About Online Abuse. *Medium* (28 May 2016) https://medium.com/humane-tech/the-immortal-myths-about-online-abuse-a156e3370aee.

Denne, Luke, Greg Sadler, and Makda Ghebreslassie. 2018. 'A Window into Your Life': Why Smart Home Devices Might Be Putting Your Privacy at Risk. *CBC*, 28 September 2018. https://www.cbc.ca/news/technology/smart-home-hack-marketplace-1.4837963.

Diakopoulos, Nicholas. 2015. Algorithmic Accountability. *Digital Journalism* 3 (3): 398–415. https://doi.org/10.1080/21670811.2014.976411.

Doctor, Vanessa. 2013. Understanding Twitter Impressions: Why Are They Important? *#Hashtags.Org*, 2013. https://www.hashtags.org/featured/understanding-twitter-impressions-why-are-they-important/.

Dwoskin, Elizabeth. 2018. A Content Moderator Says She Got PTSD While Reviewing Images Posted on Facebook. *Washington Post*, 24 September 2018. https://www.washingtonpost.com/technology/2018/09/24/content-moderator-says-she-got-ptsd-while-reviewing-images-posted-facebook/.

Eichenwald, Kurt. 2016. How Donald Trump Supporters Attack Journalists. *Newsweek*, 7 October 2016. http://www.newsweek.com/epileptogenic-pepe-video-507417.

Farokhmanesh, Megan. 2018. YouTube Is Still Restricting and Demonetizing LGBT Videos—and Adding Anti-LGBT Ads to Some. *The Verge*. 4 June 2018. https://www.theverge.com/2018/6/4/17424472/youtube-lgbt-demonetization-ads-algorithm.

Feiner, Lauren. 2019. Twitter Users Are Escaping Online Hate by Switching Profiles to Germany, Where Nazism Is Illegal. *CNBC*, 3 August 2019. https://www.cnbc.com/2019/08/02/twitter-users-switch-profiles-to-germany-to-escape-online-hate.html.

Gehl, Robert W. 2015. The Case for Alternative Social Media. *Social Media + Society* 1 (2). https://doi.org/10.1177/2056305115604338.

Geiger, R.Stuart. 2016. Bot-Based Collective Blocklists in Twitter: The Counterpublic Moderation of Harassment in a Networked Public Space. *Information, Communication & Society* 19 (6): 787–803. https://doi.org/10.1080/1369118X.2016.1153700.

Golding, Dan, and Leena Van Deventer. 2016. *Game Changers: From Minecraft to Misogyny, the Fight for the Future of Videogames*. South Melbourne, VIC: Affirm Press.

Grayson, Nathan. 2020. Twitch's Safety Advisory Council Rollout Has Been a Disaster. *Kotaku Australia* (20 May 2020) https://www.kotaku.com.au/2020/05/twitchs-safety-advisory-council-rollout-has-been-a-disaster/.

Holpuch, Amanda. 2015. Native American Activist to Sue Facebook over Site's 'Real Name' Policy. *The Guardian* (19 February 2015) http://www.theguard-

ian.com/technology/2015/feb/19/native-american-activist-facebook-lawsuit-real-name.

Horwitz, Jeff, and Deepa Seetharaman. 2020. 'Facebook Executives Shut Down Efforts to Make the Site Less Divisive'. Wall Street Journal, 26 May 2020, sec. Tech. http://archive.is/GPO6b.

House of Commons/ Home Affairs Committee. 2017. *Hate Crime: Abuse, Hate and Extremism Online*. Fourteenth Report of Session 2016–17. https://publications.parliament.uk/pa/cm201617/cmselect/cmhaff/609/609.pdf.

Jhaver, Shagun, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Online Harassment and Content Moderation: The Case of Blocklists. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25 (2): 12.

Koebler, Jason, and Joseph Cox. 2018. Here's How Facebook Is Trying to Moderate Its Two Billion Users. *Motherboard* (blog). 23 August 2018. https://motherboard.vice.com/en_us/article/xwk9zd/how-facebook-content-moderation-works.

'kovah'. 2017. Impersonation via Other Mastodon Instances Aka Fake Accounts · Issue #913. *GitHub*, 5 April 2017. https://github.com/tootsuite/mastodon/issues/913.

Lewis, Rebecca. 2018. Alternative Influence: Broadcasting the Reactionary Right on YouTube. *Data & Society*. https://datasociety.net/output/alternative-influence/.

Lomas, Natasha. 2017. Here's How to Kick Nazis off Your Twitter Right Now. *TechCrunch (blog)* (14 October 2017) http://social.techcrunch.com/2017/10/14/heres-how-to-kick-nazis-off-your-twitter-right-now/.

Martin, Allen. 2017. Twitter Can Automatically Hide Neo-Nazis and White Supremacists, but Chooses Not To. *Alphr*, 19 October 2017. http://alphr.com/go/1007424.

Massanari, Adrienne L. 2015. *#Gamergate and The Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures*. New Media & Society. https://doi.org/10.1177/1461444815608807.

Monteiro, Mike. 2018. Twitter's Great Depression. *Mike Monteiro (blog)* (17 January 2018) https://medium.com/@monteiro/twitters-great-depression-4dc394ed10f4.

Morozov, Evgeny. 2013. *To Save Everything Click Here: The Folly of Technological Solutionism*. Public Affairs.

van der Nagel, Emily, and Jordan Frith. 2015. Anonymity, Pseudonymity, and the Agency of Online Identity: Examining the Social Practices of r/Gonewild. *First Monday* 20: 3. https://doi.org/10.5210/fm.v20i3.5615.

Newton, Casey. 2020. Leaked Posts Show Facebook Employees Asking the Company to Remove Trump's Threat of Violence. *The Verge*, 29 May 2020. https://www.theverge.com/2020/5/29/21275044/facebook-trump-tweets-employee-reaction-criticism.

'@persenche'. 2020. Listen, This Is Becoming a Significant Problem. Stop. Using. Blocklists. *Twitter*, 16 January 2020. https://twitter.com/persenche/status/1217702386132643841.

Phillips, Whitney. 2014. To Fight Trolls, Focus on Actions and Context. *NY Times*, 19 August 2014. https://www.nytimes.com/roomfordebate/2014/08/19/the-war-against-online-trolls/to-fight-trolls-focus-on-actions-and-context.

———. 2015. *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture*. Cambridge, MA: MIT Press.

Pincus, Jon. 2017. Lessons (so Far) from Mastodon for Independent Social Networks. *Medium (blog)* (10 May 2017) https://medium.com/a-change-is-coming/lessons-from-mastodon-for-independent-social-networks-ae2d4ccf8f72.

Rösner, Leonie, and Nicole C. Krämer. 2016. Verbal Venting in the Social Web: Effects of Anonymity and Group Norms on Aggressive Language Use in Online Comments. *Social Media + Society* 2 (3). https://doi.org/10.1177/2056305116664220.

Rubin, Jennifer D., Lindsay Blackwell, and Terri D. Conley. 2020. Fragile Masculinity: Men, Gender, and Online Harassment. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. CHI '20. Honolulu, HI: Association for Computing Machinery. https://doi.org/10.1145/3313831.3376645.

'@_sagesharp_'. 2018. I See Some Folks Talking about Moving to Mastodon to Escape Twitter. I Don't Wanna Dash Everyone's Hopes, but You're Just Trading One Set of Problems for a Different One. Thread: Tweet. *@_sagesharp_* (blog). 16 August 2018. https://twitter.com/_sagesharp_/status/1030112338836221953.

Sarkeesian, Anita. 2015. One Week of Harassment on Twitter. *Feminist Frequency* (27 January 2015) https://feministfrequency.com/2015/01/27/one-week-of-harassment-on-twitter/.

Schipp, Debbie. 2018. Toxic Content Their 'Crack Cocaine': Facebook's Disturbing Moderator Secrets. *News.Com.Au*, 7 August 2018. https://www.news.com.au/entertainment/tv/current-affairs/toxic-content-their-crack-cocaine-facebooks-disturbing-moderator-secrets/news-story/e03358922d893e49286fe514b11fe504.

'secretgamergirl'. 2018. 'The New McCarthyism'. Secret Gamer Girl. 1 April 2018. https://secretgamergirl.tumblr.com/post/172476575235/the-new-mccarthyism.

Shepherd, Tamara, Alison Harvey, Tim Jordan, Sam Srauy, and Kate Miltner. 2015. Histories of Hating. *Social Media + Society* 1 (2). https://doi.org/10.1177/2056305115603997.

Silverman, Robert. 2016. Twitter Is Deleting Olympics Videos. Harassment? Nah. *Vocativ*, 15 August 2016. http://www.vocativ.com/350674/twitter-is-deleting-olympics-videos-harassment-nah/.

'sjwomble'. 2016. The Problem with Personal Block Lists. *A Blog* (blog). 28 April 2016. https://sjwomble.wordpress.com/2016/04/28/the-problem-with-personal-block-lists/.

'@th3j35t3r'. 2017. 'CounterSocial'. Counter.Social. 2017. https://counter.social/about.

———'. 2018. Pakistan Allowed OBL to Hole up next to One of Their Military Bases for 7 Years (at Least). So... Twitter (Archived). *@th3j35t3r* (blog). 16 August 2018. https://web.archive.org/web/20181012045938/https:/twitter.com/th3j35t3r/statuses/1030281783034576897.

Thompson, Peter A. 2019. Beware of Geeks Bearing Gifts: Assessing the Regulatory Response to the Christchurch Call. *The Political Economy of Communication;* Vol 7, No 1 (2019), August. http://www.polecom.org/index.php/polecom/article/view/105/314.

Tilley, Sean. 2017. A Quick Guide to The Free Network. *We Distribute (blog)* (23 September 2017) https://medium.com/we-distribute/a-quick-guide-to-the-free-network-c069309f334.

Tufekci, Zeynep. 2018. Opinion | YouTube, the Great Radicalizer. *The New York Times*, 10 March 2018, sec. Opinion. https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.

Turner, Fred. 2006. *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago; London: University of Chicago Press.

Ullrich, Johannes. 2012. InfoSec Handlers Diary Blog - An Analysis of Jester's QR Code Attack. (Guest Diary). SANS Internet Storm Center. 2012. https://isc.sans.edu/diary.html.

Vance, Brendan. 2014. Form and Its Usurpers. *Brendan Vance (blog)* (17 July 2014) http://blog.brendanvance.com/2014/07/16/usurpers/.

Vella, Heidi. 2018. IoT Devices and Smart Domestic Abuse: Who Has the Controls? *Engineering & Technology*, 20 June 2018. https://eandt.theiet.org/content/articles/2018/06/iot-devices-and-smart-domestic-abuse-who-has-the-controls/.

Wagenseil, Paul. 2012. Anti-Anonymous Hacker Threatens to Expose Them. *Msnbc.Com*, 13 March 2012. http://www.nbcnews.com/id/46716942/ns/technology_and_science-security/t/anti-anonymous-hacker-threatens-expose-them/.

Warzel, Charlie. 2016a. 'It Only Adds To The Humiliation'—How Twitter Responds to Harassers. *BuzzFeed News*, 22 September 2016. https://www.buzzfeednews.com/article/charliewarzel/after-reporting-abuse-many-twitter-users-hear-silence-or-wor.

———. 2016b. 90% Of The People Who Took BuzzFeed News' Survey Say Twitter Didn't Do Anything When They Reported Abuse. *BuzzFeed*, 23 September 2016. https://www.buzzfeed.com/charliewarzel/90-of-the-people-who-took-buzzfeed-news-survey-say-twitter-d.

West, Lindy. 2014. Twitter Doesn't Think These Rape and Death Threats Are Harassment. *Daily Dot* (23 December 2014) http://www.dailydot.com/opinion/twitter-harassment-rape-death-threat-report/.

Whigham, Nick. 2018. Film Shows the Unlikely People Tasked with Cleaning up Social Media and the Ugly Consequences. *News.Com.Au*, 14 October 2018. https://www.news.com.au/technology/online/social/film-shows-the-unlikely-people-tasked-with-cleaning-up-social-media-and-the-ugly-consequences/news-story/e8eb561b983b9a13e5a26f8844bddaf0?from=rss-basic.

Yeo, Amanda. 2020. 'One Twitter Account Is Reposting Everything Trump Tweets. It Was Suspended within 3 Days.' Mashable, 3 June 2020. https://mashable.com/article/twitter-donald-trump-suspend-tweets-policy-violence/.

CHAPTER 6

Reshaping the Landscape

The last chapter discussed tools frequently proposed to solve online harassment, but which instead either do not work or make the situation worse, alongside financial motivations online platforms have not to change the status quo. This chapter proposes methods that could be used to limit online harassment and mitigate its impact by approaching harassment communities as alternate reality games.

Although, as discussed in Chap. 3, it is functionally impossible to shut down a harassment community from the outside, it is possible to set up online spaces—and thus networked publics—that discourage the formation of harassment communities and make users safer from abuse.

The central principle of designing networked publics in order to make them inhospitable for harassment communities is to put agency in the hands of the people who use them and allow them as individuals to control their own engagement with the online space. Harassment communities depend on the affordances of the networked public to create areas where the users do not have control over their engagement and essentially trap them against those walls: if you want to use the space, you have to accept this dimension to your engagement. Twitter is an obvious example: there are very limited tools for filtering your mentions or replies, meaning that if a harassment community wants to dogpile you, it is hard to stop them.[1] In comparison, Rosie Pringle highlights the significant impact that giving people agency over their experiences of engaging with a networked public can have (Pringle 2016). Pringle argues that although a set of

K. Veale, *Gaming the Dynamics of Online Harassment*,
https://doi.org/10.1007/978-3-030-60410-3_6

external rules can be gamed, it is much harder to circumvent 'rules' applied at an individual level—and it is easy to correct if someone tries (Pringle 2016).[2]

In this chapter I will explore a series of possibilities that a networked public could use to raise the level of agency of its users and thus make the pragmatic labour of harassment communities harder and less satisfying to accomplish. Many of these suggestions were never formally published or originate from authors who asked not to be cited due to the likelihood of receiving more harassment if they were noted here: it is both safe and important to assume that they originate from a broader community rather than myself if they are left without specific citation.

All of these possibilities would be best explored through methodologies of intersectionally feminist Human-Computer Interaction (HCI) to ensure that they contribute usefully to safe and inclusive networked publics (Bardzell 2010; Bardzell and Bardzell 2011; Fiesler et al. 2016; Suchman 2009). A vital part of that process would be to have as many diverse perspectives on them as possible from conceptualisation, implementation and through to development, including asking people the seemingly simple question, 'how can these be used by malevolent actors?' Refinement after release based on feedback would also be necessary, given how creatively unpredictable bad actors can be.

Filtering and Visibility

One example of a change to networked publics that would dramatically limit the ability of harassment communities to control the experiences of the people they target would be to design infrastructure to offer easy filters for replies, mentions and so on. If users had access to a channel or stream entirely made up of communication from known friends, then they could continue to communicate with those people unhindered by a flood of harassment elsewhere (Auerbach 2016; Harper 2016). Not everyone using the platform would require this option, so allowing people to select a mixed list of all communication, a filtered list of communication from friends, a filtered list of communication from strangers, or potentially as many as desired in parallel, would put control back into the hands of people using a given platform. In 2017, Twitter released a series of optional filters for notifications to improve this facet of the service (Cohen 2017; Shaul 2017). However, in 2020, the options remain largely

unpopularised, and some users are surprised to discover they exist after complaining of their absence.

Another important form of opening up control would be to replace the binary decision to have your posts open to the public or be closed, as currently exists on sites like Twitter, which poses a significant restriction on an individual's ability to use social media. Instead, users should be allowed to set variable levels of access/engagement—and in parallel to access, users need to be able to set variable levels of *visibility*. Control over visibility would mean that a given users' content simply cannot appear to individuals within categories they set, even if reshared by third parties or searched for directly. Blocking access from accounts in key categories limits their capacity to reshare posted material, to reply and so on, but means the content can still be seen. Rendering content invisible to people in key categories goes further in that it means that the content functionally does not exist for them and cannot be engaged with at all—essentially meaning that a harassment community would be unaware a given person was using the platform. An obvious option to provide users with here would be to control whether people who are not logged into the platform can access your content, or even see it at all: if applied, these options would automatically prevent block-evasion as currently exists on most platforms, because anyone logging out to avoid a block would still be unable to reach their target. The option of setting individual posts or potentially entire accounts to deny engagement or be invisible to anyone *except* lists of specific named accounts would be useful for controlling context collapse, as has been possible using Google Plus, LiveJournal and Dreamwidth.

Danilo Campos proposes a number of categories that users should be able to block, and I argue that they would be equally valid as criteria to become invisible to as well. Campos proposes optional filtering based on

- the age of the accounts;
- the number of people the accounts are following;
- whether the accounts use particular words/phrases/hashtags nominated by the user and
- whether the accounts are blocked by more than a given number of people the user is following[3] (Campos 2014).[4]

Users should be able to access spectrums of filtering on each axis by being able to specify the points at which each triggers, because that puts agency back in their hands.[5] Pringle likewise proposes the ability to mute/filter

any posts based on any content the user chooses not to see—and suggests that the ability to share these lists between users would be useful if it were done in a transparent fashion (Pringle 2016). Caroline Sinders proposes the option of filtering accounts that you have no friends in common with (Sinders 2015). Campos highlights that blocks need to be 'opaque,' in that people who have been restricted need to have no sign that they have been filtered, and this is potentially another area where the option to control opacity should be provided to users (Campos 2014). The risks associated with blocking being visible[6] are reduced if the platform features options which guard against block-evasion, and individual users will be able to assess how controlling the visibility of blocks would affect their engagement with the networked public.

Further possibilities for filtering could tie to layered, parallel authentication options for accounts, such as where people could tie an account to a phone number, credit card and/or physical address.[7,8] Providing multiple options would make some level of authentication flexibly available to as many people as possible as a way to help screen out bot accounts. Allowing users to block or become invisible to unauthenticated accounts (potentially allowing them to select which authentication methods—or combination of methods—to allow) would be another means by which they could curate and control their experience of a space. A further benefit of such a system is that it would add to the existing filtering systems to create concrete consequences for poor behaviour on the network. If you are blocked by someone, in the system proposed here you cannot block-evade simply by logging out, and although you could potentially create a new unauthenticated account quickly, the flexible filtering options mean there is no guarantee you will be visible to (or even able to see) anyone who just blocked you. If there were visible cues of authentication such as colour tags or badges signalling the different kinds (or multiple layers) that have been applied, this would potentially make bot-based amplification of messages more visibly unreliable in comparison to human accounts. If authentication is flexible, secure, simple to apply and visibly communicated, then accounts lacking any authentication methods at all will likely stand out as unusual.

Giving users control over their content would be a vital supplement to these flexible filtering systems, such as allowing users to control how individuals or audiences can engage with their content. For example, users ideally need to have modular options at the post level allowing them to create public posts that are visible but cannot be responded to or reshared,

all the way through to posts that can be reshared and responded to but which will only be visible to users fitting key criteria. In addition to being able to filter who can see/interact with content even once it has been reshared, users need to be able to delete that content and have that deletion carry through the network. Ideally, users also need to be able to change the visibility/engagement settings on posts after the fact, as they adjust the level of energy they have to handle responses, or if they get unexpected, unwanted attention from harassment communities.

Third-Party Reporting

Allowing third-party reporting would mean that people targeted by harassment communities are not expected to handle the problem themselves (Coccimiglio 2015). Studies have shown that less than half of reports of harassment or abuse come from the people attacked themselves, highlighting how useful an option it is to provide (Geiger 2016, 792; Kessler 2015). One advantage of third-party reporting being a possibility is that it provides a secondary line of response that does not assume the flexible filtering systems proposed here will innately solve harassment problems. The reporting process also needs to be as streamlined and intuitive as possible to encourage people to engage with it—and that includes closing the loop by noting whether action was taken, even if the specifics remain confidential. Contextual information regarding the person making the report and their relationship to the person they are reporting attacks on would also be useful: if they are reporting a reply to a post, are they following the person being replied to? Is the person making the reply also following the person who they are responding to, or are they otherwise not connected? Is there topical bot activity related to them or to events they have discussed prior to the report (Lapowsky 2018)? There are layers of information that could help diagnose mass-reporting, even if the flexible filtration systems would make it less likely: such a system provides parallel tools for dealing with mass-reporting by hiding content and/or accounts from attackers, without preventing or limiting users from engaging with the affordances of the social network.

Human Moderation, Clarity of Site Rules and Precision in Tailoring Individual Experiences

The layers of information associated with reports need to be sent to human agents rather than automated processes. Those humans need to be able to respond without any restrictive rules limiting their ability to act against hate speech of the kind that Facebook provides (Koebler and Cox 2018). There is no algorithmic solution to online harassment, and human oversight is vital since context always matters. However, since it is both vital and a dangerous job, all moderation staff need to be provided with mental health support for dealing with horrific and traumatising content (Dwoskin 2018; Whigham 2018). In addition, there needs to be transparency around abuse on the platform, such as numbers of account suspensions and content takedowns related to abuse complaints, the numbers of complaints, type of complaint and what the results of complaints were. Without reliable data on abuse, it is difficult to hold platforms accountable for their response to it or to trust their public-facing statements on the subject (Coccimiglio 2015; Dewey 2014).

Rules against impersonation at all levels—such as creating fake accounts, editing posts or screenshots of posts, and distributing material from the above, along with other techniques—need to explicitly be against the rules of the platform. There need to be reporting options associated with them, along with similar rules against posting personal information for the purposes of harassment (Alexander 2016).

Randi Harper lists a series of specific technical suggestions for modifying Twitter's infrastructure to improve peoples' ability to deal with and respond to harassment, and they would be a good checklist for the affordances of other platforms as well (Harper 2016). Muting all replies to a specific post needs to be an option, along with blocking or muting a hashtag or equivalent form of organising structure. Users need to be able to auto-mute any replies that mention your account made to users who you have blocked. Using the platform's 'search' function needs to have options for filtering and excluding content from blocked users, and for hiding your content from people fitting categories you specify.

All of these specific solutions would likely be less critical in a context with functional, flexible filtering and visibility settings. However, in some ways they function as technological canaries for the platform's coal-mine: if a user wanted to handle a specific problem along these lines, would the platform give them the tools they need for tailoring their experience?

Additionally, Harper argues that it needs to be possible to list posts and content that has otherwise been directly blocked or filtered, and *only* those, to facilitate evidence-gathering and reporting of the accounts that are a problem (Harper 2016): if users lack these options, pursuing their harassers through formal complaints or the legal system is difficult. Likewise, there needs to be a list of content you have reported, ideally that ties to formal actions from the platform made in response.

All of these settings need to be remembered by the platform itself so that there is as little repetition or friction for the user as possible in setting the boundaries and circumstances through which they intend to engage with the networked public. Making something possible but requiring the user to re-establish their parameters each time they log in, or on some other schedule, functionally establishes a default for the site that is outside their control, rather than supporting their agency across their experience of the networked public.

Putting as much flexible, individually tailorable control as possible into the hands of users defangs a significant proportion of existing harassment campaign tactics and makes it harder for the tiers that make up harassment communities to operate. If they cannot see or engage with accounts that have filtered them, they cannot get information out of those target accounts back to the harassment community, and they simultaneously cannot easily reach targets with their harassment. If their harassment does successfully reach target accounts, then those accounts have multiple options for responding to the problem—and some foundational design discussed here would limit how much that harassment prevented the targets from being able to use the site.

One tactic members of harassment communities employ which would survive the methods suggested here is to follow someone with an entirely innocent-seeming account, using it to pass information to the community. Although it is possible that surveillance-accounts could survive individually tailorable filters and maintain contact with the person being targeted, the tools would make responding to the problem easier. Firstly, new attempts to 'innocently' surveillance-follow someone already concerned about harassment would be filtered by existing controls on how old an account needs to be before it can see or potentially interact with their posts or account. Additionally, if someone discovered that their information *was* being passed to a harassment community despite their filters, they would have the ability to lock down further and test at what point the problem stopped. Although this process could only be done by the person

being harassed, the ability to experiment in order to isolate probable problem accounts would be a possibility in a way it otherwise would not.

Disrupting the Tiers of Harassment Communities

All of the features discussed here contribute towards disrupting the ability of tiers within the harassment community to function. They limit the capacity of the different tiers to gather information and their ability to communicate with their targets. In turn, those limitations will reduce their impact and even their motivation to form on the platform.

A problem is that if a harassment community cannot function on a given platform, they will attack on platforms better suited to the purpose. The upside is that even if this were the case, there would be a safer space online for people being targeted by harassment campaigns to exist within, and all of the harassment community strategies tied to isolating targets and destroying their ability to communicate and ask for help would no longer work.

If the platform had sensible rules, reporting tools and responses to harassing behaviour, that would disrupt the ability of tiers to communicate internally among themselves. Even in absence of formal responses to harassing behaviour, flexible filtering tools tied to visibility and engagement would create consequences for poor behaviour because it would no longer be possible to evade blocks and return with newly formed accounts. As a result, these approaches avoid many of the problems associated with offloading the effort and responsibility of solving the problem of harassment onto already marginalised people. Those marginalised people would have significant ability to control their engagement with the networked public via these tools, without needing to rely solely on formal support from the platform.[9]

The goal would be to change the status quo where defending against abuse is more time and energy than abusing someone, into a dynamic where users have tools which let them reduce the impact of abuse easily, potentially ahead of time. Simultaneously, the effort required to reach someone and harass them would rise. That is not a defence in itself, since people in harassment communities have shown a terrifyingly dedicated willingness to spend incredible time and energy to the cause of abuse, but every little bit helps. Every incremental change which makes abuse more difficult or easier to evade will likely reduce the size of the tertiary, audience tier of the harassment community. Reducing the size of the tertiary

tier will correspondingly reduce the number of people who then choose to rise into the secondary and primary tiers. It will also reduce the number of people cheerleading for the existing people in those more active tiers.

It is a massive understatement to say that the suite of options suggested in this chapter would pose challenges for conceptualisation, design and implementation. However, the key is that the more of these principles any given platform can achieve, the better, and most existing platforms are not achieving many of them—if any. Regardless of the specifics for how doing so is accomplished, the central principle of designing online spaces and the networked publics they shape needs to become making them as hard to weaponise as possible and as easy for users to control their engagement with. Any platform which does less is directly complicit in the abuse committed through their networked publics, either due to negligence or via profiting from it.

Case Study: Pillowfort.Social

Pillowfort.Social is an example of alternative social media (ASM). It does not gather and on-sell information from its users to advertisers, nor does the site present advertising to users as part of the site design unless they explicitly consent to them. Instead, its business model follows Dreamwidth in approaching the users as potential customers who can be sold expanded services. The site began development in 2015, eventually raised US$57,000 development funding through Kickstarter in 2018 and then moved to a model where accounts could be purchased for US$5 each during a closed beta ('PretzelSpaceship' 2019). Since November 2019, users can share up to three invite codes a week for free, and the intent is for the site to open to the public in 2020.

Pillowfort is notable for the ways that its underlying design has been constructed to produce a networked public that is relatively resistant to online harassment.[10] For example, the site features an accessibly written policy document (Baritz n.d.) and privacy policy (Baritz 2019), and those documents explicitly define behaviours associated with online harassment as problems that can be reported and will be acted on. The site's Twitter account explicitly mentions banning neo-Nazi accounts ('@pillowfort_soc' 2018), suggesting that the policies are tied to active practices of supporting the community.

A central pillar of Pillowfort's design is to give the users control over their posts and how those posts are engaged with. For example, each post can be

- made public
- restricted to be visible only to people logged into the site;
- restricted to be visible only to people following the user's account[11];
- restricted to be visible only to people following the user who the user is also following back, a category called 'mutuals' or
- restricted to the user themselves.

In addition, posts can be restricted so that nobody can comment on them or nobody can reshare them. Restricting the audience who can see a post marks it as non-rebloggable by default, but this can be turned on again by the user.

These settings provide a granularity of visibility and engagement, controlled by the user, and importantly, this can be changed at any point: there is effectively only one copy of each post, and that copy remains under the author's control at all times. As a result, if they edit a given post, delete it or change any of its visibility or engagement settings, those changes propagate through Pillowfort's network. This avoids the problem 'Coyote' identified where Tumblr's design created innate context collapse, and guaranteed that authors lost control of posts as soon as they are reshared, as discussed in Chap. 4 ('Coyote' 2019).

Pillowfort has threaded replies in response to posts, something which contributes to substantial changes compared to other networked publics. Firstly, this reinforces that the authors of posts retain control over that 'space' within the networked public, because all engagement with their posts happens where they can see and control it. Even if a post is reshared into a community that the author does not participate in, they will see responses to it. Secondly, it avoids another of 'Coyotes' critiques of Tumblr, in that since conversations are encouraged by the site design, there is no need to reshare material in order to critique it, changing the tone of the networked public overall ('Coyote' 2019).

When there are problems in the responses to a given post, users are given agency in responding to them. The problem replies can be deleted without fuss, and if a given poster is causing consistent problems, blocking them is also simple. The blocking tools are robust and instantly render both the person blocked and the person who blocked them invisible to

each other, meaning that they cannot encounter each other's material again even by accident.

In the case of surveillance-accounts, Pillowfort presents easily accessible lists of who has followed your account and for how long. This is not a perfect defence by any means, but would help with narrowing down the likely offender.

Users also have access to blacklist filters which give them agency over rendering posts using particular keywords as tags, or in tags as well as the body of posts, completely invisible to them.

There are ways that Pillowfort could improve the design of its networked public to give people more control over their experiences in it.[12] For example, although there are granular controls at a post level, each post defaults to being unrestricted: allowing users to set up personalised defaults would streamline the process. Also, controls over post visibility could be expanded: currently, there is no way of restricting based on the age of the account, for example, which would be useful for dealing with someone who has been blocked immediately creating a new account. There is also no current ability to post in a way that is visible only to a particular person or collection of people, for those situations where it is not exclusive enough to restrict a post to everyone you are following who is also following you. It would also be useful to increase the number of post flags available, expanding the current 'Not Safe for Work' flag to exist alongside an 'Adult Content' flag and potentially others. Using one flag to cover both situations causes problems, discourages some people from correctly applying flags and means people who would like to filter one option but not others either encounter material they would prefer not to or cut out content they would otherwise like to see.

Pillowfort.Social's underlying design and conceptualisation provides agency to the people using the site, with substantial consequences for the networked public growing up within it. Although there are ways its design could be improved further, the fundamental philosophy of approach puts the experience of users as its focus, rather than its ability to absorb information from them or otherwise monetise them. In this regard, it is a good example of the strengths Gehl identifies as being possible through ASM (Gehl 2015). One of the results of providing its users with agency over their own posts and experiences of the site is that they would be substantially more difficult to harass in this context and have more tools for responding if someone makes the attempt. It also seems likely that the site itself would be more proactive than many other platforms have been in

responding to problems. One of the reasons for this is that the site's business model is grounded in attracting and keeping an audience who enjoy the space and who can be sold expanded account options. Making a pleasant space people want to spend time in *is* their business model.

Case Study: Ahwaa

In contrast with Pillowfort.Social, Ahwaa has designed its networked public to be as safe as possible through requiring good citizenship in order to participate. Ahwaa is an Arabic-language forum for LGBTQA+ people founded by Esra'a Al Shafei as an environment where people can be safe to discuss rainbow issues and meet each other (D'Arcy 2018). In order to protect the networked public from both ubiquitous internet bigots and police forces targeting LGBTQA+ communities, Ahwaa is constructed as a series of nested spaces arranged so that new members only have access to the outermost shell. The system works to use gamification to encourage good citizenship, and the site does not make people aware their access is limited: it simply opens up further as they contribute to the community:

> Every positive interaction—such as sharing an experience, giving advice, making a comment or asking a question—earns points for a user. Users can also award points to others based on the quality of their interactions with that person. For instance, if you create a topic or thank someone for their post or thread, you receive five points. If you leave a comment that a user finds "helpful", you get ten. This encourages people to be more responsive and friendly and to leave comments with substance—as Al Shafei puts it, "you can visit once, leave a super-insightful comment, and earn hundreds of points." Users need 100 points to progress to the first level, 500 for the second, and 2,000 for third. (…) Most trolls on Ahwaa never make it past the very first chat rooms. "If there was someone trying to infiltrate the system, they would have spend many hours being super supportive and cooperative," Al Shafei says. "To really win against the trolls, we've found you have to exhaust them." (D'Arcy 2018)

Ahwaa has been operating for over a decade in which bad actors have never reached the more private spaces. Al Shafei ascribes this success to the site developing a supportive community invested in quality engagement, which has produced insulation from malevolent infiltration. Additionally, Al Shafei's point that Ahwaa is more exhausting for bad actors to try and infiltrate than it is for users to simply exist and communicate within is vital

(D'Arcy 2018) and stands in a stark contrast to most social media platforms, where it is far easier to attack people than it is to defend yourself.

Chapters 3 and 4 have explored why the current status quo is intolerable and puts marginalised groups at constant risk of life-ruining harassment. Chapter 5 argued that the status quo is not neutral and has happened because it is profitable to companies who are either indifferent to harassment or who actually profit from it directly.

This chapter has illustrated that alternatives to the status quo are possible: that the underlying design of networked publics can have a positive impact on creating spaces which resist online harassment, and are safer for people from marginalised groups as a result. The key is to provide people with as much flexible control over their own engagement with a space as is possible since harassment campaigns depend on removing agency or finding places where the design of the networked public removes it for them.

However, we are left with the problem of how to make that happen. It is clear that the status quo is not going to change by itself, and so the next and final chapter explores what the options may be for what will happen next.

Notes

1. Yishan Wong lists a collection of traits that add up to how easy Twitter is to abuse (Wong 2016).
2. Pringle's comment is specifically in the context of taking inspiration for social media design from Massively Multiplayer Online videogames (MMOs) and on the design of profanity filters. However, I argue the point applies more broadly.
3. I suggest that blocks derived from an account being blocked by more than a given number of people a user is following should only apply if they have been blocked *directly*. That would avoid the possibility of contagion where blocks/filtering are automatically applied, raising the number of people blocking the account, which then raises the odds that they will be automatically blocked by other users applying the option. Potentially letting users decide whether the blocks are applied directly or whether they are comfortable with the possibility of contagion would be another layer of control for this option.
4. Many of these possibilities have been introduced as optional filters for Twitter in 2017, but as noted earlier, their existence has not been advertised and a large proportion of users are unaware they exist.

5. As noted, Twitter introduced some of these concepts in 2017, but they are binary options where the user cannot adjust their own preferences for each, and things like what defines a 'new' account are opaque to the system.
6. Such as blocks being seen as a sign of 'winning' against a target, something harassers compete to collect, or a cue to try circumventing the security of the person being targeted.
7. Encrypted random number generating apps might be a possibility as well since they would be useful for confirming a level of human presence: the downside would be that the people who create and organise networks of bots could likely apply such an authentication method to their armies of fake accounts with fewer hurdles than the other options here.
8. It would be important that this information not be accessible or ideally even stored on social network servers wherever possible, to reduce the incentive to try and steal the relevant databases. Online auction sites have been handling problems like these by generating a random number code and sending it to an address via physical mail to authenticate a given address, without necessarily storing the address in their systems, so there are workarounds.
9. Ensuring that there was also formal support *in addition to* the ability to curate and control their own online spaces would be vital. However, it seems safe to suggest that a networked public defined by user agency and flexible tools *without* formal support and responses to reports would likely be safer than a space with good formal responses but which lacked those tools. Any platform is going to have multiple audiences with different needs, and so focusing on reporting systems and formal responses is certainly an improvement over many existing platforms. Unfortunately, this approach in isolation raises the chances that harassment campaigns will find areas where some parts of the user base *can* be attacked more easily because they cannot control their engagement with the space in the ways they need.
10. Given that the site attracted the attention of both Breitbart and Kiwi Farms in 2015 as being worthy of harassment, this may be as much a pragmatic response to the reality of online harassment as it is a foundational design philosophy in itself (Bokhari 2015; 'JU 99' 2015).
11. This option would neatly thwart attempts to block-evade by logging out the blocked account or viewing by browser.
12. There are also accessibility options that could be expanded across the site, to better allow people to shape their experiences around their own needs. For example, there is no way to change the aesthetic design of the main feed of posts on the site, and this causes problems for some users with vision problems. There are reasonably robust options for changing how *your part of the site is seen by other people*, but less how you see the rest of the site.

References

'@pillowfort_soc'. 2018. Even Better: We Ban Nazis from Our Platform without Being Silicon Valley! *Twitter*, 9 August 2018. https://twitter.com/pillowfort_soc/status/1027285350077616128.

Alexander, Leigh. 2016. Milo Yiannopoulos: Twitter Banning One Man Won't Undo His Poisonous Legacy. *The Guardian* (20 July 2016) https://www.theguardian.com/technology/2016/jul/20/milo-yiannopoulos-twitter-ban-leslie-jones-bad-idea.

Auerbach, David. 2016. Twitter Needs a Drastic Plan to Save Itself. Here It Is. *Slate.Com*. 25 January 2016. http://www.slate.com/articles/technology/bitwise/2016/01/twitter_needs_a_drastic_plan_to_save_itself_here_it_is.html.

Bardzell, Shaowen. 2010. *Feminist HCI: Taking Stock and Outlining an Agenda for Design*. 1301–1310. ACM.

Bardzell, Shaowen, and Jeffrey Bardzell. 2011. *Towards a Feminist HCI Methodology: Social Science, Feminism, and HCI*. 675–684. ACM.

Baritz, Julia. 2019. Privacy Policy. *Pillowfort.Social*, 24 March 2019. https://www.pillowfort.social/privacypolicy.

———. n.d. Terms of Service. *Pillowfort.Social*, Accessed 21 April 2020. https://www.pillowfort.social/TermsOfService.

Bokhari, Allum. 2015. En Route: A Site for People Who Find Tumblr Too Stressful. *Breitbart* (8 November 2015) https://web.archive.org/web/20200528033047/https://www.breitbart.com/tech/2015/11/08/theres-a-tumblr-for-people-who-find-tumblr-too-stressful-and-hostile/.

Campos, Danilo. 2014. The Least Twitter Could Do. *DaniloCampos.Com*, 29 July 2014. https://write.danilocampos.com/the-least-twitter-could-do-6d77ba1ccc05.

Coccimiglio, Jessica. 2015. Tweeting While Female: Harassment, and How Twitter Can Fix It. 30 March 2015. http://www.makeuseof.com/tag/abuse-twitter-happens-twitter-can-fix/.

Cohen, David. 2017. Twitter Just Added Several Advanced Filters So Users Can Control Their Notifications. *Adweek*, 10 July 2017. https://www.adweek.com/digital/twitter-advanced-filters-notifications/.

'Coyote'. 2019. How Using Tumblr Is Undermining Your Community. *The Ace Theist* (blog). 23 November 2019. https://theacetheist.wordpress.com/2019/11/23/how-using-tumblr-is-undermining-your-community/.

D'Arcy, Patrick. 2018. The Smart Strategy That One LGBTQ Forum Uses to Keep out Trolls and Bullies. *Ideas.Ted.Com* (blog). 21 September 2018. https://ideas.ted.com/the-smart-strategy-that-one-lgbtq-forum-uses-to-keep-out-trolls-and-bullies/.

Dewey, Caitlin. 2014. New Report Slams Twitter, Facebook and YouTube for Secrecy around Harassment of Women Online. *Washington Post*, 16 September 2014. https://www.washingtonpost.com/news/the-intersect/wp/2014/09/16/new-report-slams-twitter-facebook-and-youtube-for-secrecy-around-harassment-of-women-online/.

Dwoskin, Elizabeth. 2018. A Content Moderator Says She Got PTSD While Reviewing Images Posted on Facebook. *Washington Post*, 24 September 2018. https://www.washingtonpost.com/technology/2018/09/24/content-moderator-says-she-got-ptsd-while-reviewing-images-posted-facebook/.

Fiesler, Casey, Shannon Morrison, and Amy S. Bruckman. 2016. An Archive of Their Own: A Case Study of Feminist HCI and Values in Design. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2574–2585. San Jose, CA, USA: ACM. https://cfiesler.files.wordpress.com/2016/02/chi2016_ao3_fiesler.pdf.

Gehl, Robert W. 2015. The Case for Alternative Social Media. *Social Media + Society* 1 (2). https://doi.org/10.1177/2056305115604338.

Geiger, R.Stuart. 2016. Bot-Based Collective Blocklists in Twitter: The Counterpublic Moderation of Harassment in a Networked Public Space. *Information, Communication & Society* 19 (6): 787–803. https://doi.org/10.1080/1369118X.2016.1153700.

Harper, Randi. 2016. Putting out the Twitter Trashfire. *Medium*, 13 February 2016. https://medium.com/art-marketing/putting-out-the-twitter-trashfire-3ac6cb1af3e#.48vatomgo.

'JU 99'. 2015. Pillowfort. *Kiwi Farms*. 9 November 2015. https://web.archive.org/web/20200528033159/https://kiwifarms.net/threads/pillowfort.14342/.

Kessler, Sarah. 2015. A Snapshot Of How Twitter Deals With Online Harassment. *Fast Company* (13 May 2015) https://www.fastcompany.com/3046262/a-snapshot-of-how-twitter-deals-with-online-harassment.

Koebler, Jason, and Joseph Cox. 2018. Here's How Facebook Is Trying to Moderate Its Two Billion Users. *Motherboard* (blog). 23 August 2018. https://motherboard.vice.com/en_us/article/xwk9zd/how-facebook-content-moderation-works.

Lapowsky, Issie. 2018. Here's How Much Bots Drive Conversation During News Events. *Wired*, 30 October 2018. https://www.wired.com/story/new-tool-shows-how-bots-drive-conversation-for-news-events/.

'PretzelSpaceship'. 2019. An Abridged History of Pillowfort Alpha & Beta. *Pillowfort.Social*, 1 April 2019. https://www.pillowfort.social/posts/563380.

Pringle, Rosie. 2016. What World of Warcraft Can Teach Us about Twitter Harassment. *Medium*, 13 January 2016. https://medium.com/@mostlyoriginal/what-world-of-warcraft-can-teach-us-about-twitter-harassment-164a9d3e2ec8#.rxzxd7fu6.

Shaul, Brandy. 2017. Twitter Introduces New Notifications Filters, More. *Adweek*, 1 March 2017. https://www.adweek.com/digital/twitter-introduces-new-notifications-filters-more/.

Sinders, Caroline. 2015. That Time the Internet Sent a SWAT Team to My Mom's House. *Narrative.Ly* (17 July 2015) http://narrative.ly/that-time-the-internet-sent-a-swat-team-to-my-moms-house/.

Suchman, Lucy. 2009. Agencies in Technology Design: Feminist Reconfigurations. *In Proceedings of 5th European Symposium on Gender & ICT, Digital Cultures: Participation–Empowerment–Diversity.*

Whigham, Nick. 2018. Film Shows the Unlikely People Tasked with Cleaning up Social Media and the Ugly Consequences. *News.Com.Au*, 14 October 2018. https://www.news.com.au/technology/online/social/film-shows-the-unlikely-people-tasked-with-cleaning-up-social-media-and-the-ugly-consequences/news-story/e8eb561b983b9a13e5a26f8844bddaf0?from=rss-basic.

Wong, Yishan. 2016. What Does Twitter Need to Do to Restart Growth and Reactivate Its Massive Dormant User Base? *Quora*, 12 February 2016. https://www.quora.com/What-does-Twitter-need-to-do-to-restart-growth-and-reactivate-its-massive-dormant-user-base/answer/Yishan-Wong?srid=XkXe&share=2a0359e7.

CHAPTER 7

Conclusion: The Christchurch Call to Action Summit and What Follows

The vast volumes of online harassment which flows through social networks and online spaces are not a new problem and disproportionately target members of marginalised groups. Online harassment is the visible and digital expression of an ongoing culture of abuse targeting women and other marginalised groups across global societies, grounded in the entitlement and insecurity of cisgender, straight, white men.

The design of online contexts shapes the networked publics that form within them: in particular, internet services and social networks have unspoken or hard-coded rules that either do not prevent abuse or can be turned into weapons in the service of abuse.

Harassment campaigns are best understood as autonomous alternate reality games (ARGs) shaped by the networked publics they operate within, which frame terrorising their targets as the goal of the game and which award social and literal capital to the individuals who contribute best to that goal. The fact these cybermobs display analogous internal community structures and patterns of affective investment as ARGs suggests that we can use this comparison to explore ways of discouraging them from forming and functioning in the future.

Providing users tools and agency with which to control and tailor their own experience of online spaces would disrupt the community dynamics and strategies that harassment campaigns function through, because they rely on online spaces taking that control away.

K. Veale, *Gaming the Dynamics of Online Harassment*,
https://doi.org/10.1007/978-3-030-60410-3_7

Terraforming social networks and the networked publics they mediate to be more hospitable to online citizens than they are for organised harassment campaigns weaponising them means we can protect the ability of people—particularly marginalised people—to exist online.

However, a fundamental problem that will need to be overcome is that existing social networks and online spaces are either indifferent to, or profit from, abuse committed via their networks. Alongside the *algorithmic accountability* suggested by Nicholas Diakopoulos, there is an equal need for *affordance accountability* that assesses how the affordances of online spaces impact the networked publics that they mediate (Diakopoulos 2015). This would help highlight where social networks are abandoning their responsibilities, and either inspire better design of new, safer online spaces or encourage existing online spaces to change.

Given that, observably, the free market has not provided a solution for online harassment despite the fact that alternatives exist, we are left with the problem of how to make change happen. The status quo is intolerable, so what next?

The Christchurch Call to Action Summit is an attempt to change the global conversations around the status quo and one which makes some concrete positive contributions alongside some troubling elements and areas that introduce more problems than they solve.

Contextualising the Christchurch Call to Action Summit

In direct response to the white-supremacist terrorist attacks against the Christchurch Al-Noor Mosque and Linwood Avenue Islamic Centre in Aotearoa-New Zealand in March 2019, the government of Aotearoa announced the Christchurch Call to Action Summit (The Call) on the 24th of April that year (Ardern 2019a). Spearheaded by Aotearoa's Prime Minister Jacinda Ardern and Emmanuel Macron, Prime Minister of France, the summit occurred in Paris on 15 May 2019. The purpose of the summit was to begin a global discussion regarding ways to combat the proliferation of violent extremist content online (Ardern 2019b), and brought together a coalition of nations[1] and large corporations.[2] The pledge document that signatories nominally agree to is non-binding, is three-pages long and contains provisions under three broad umbrellas, all of which were published publicly online ('Christchurch Call' 2019).

The Good

The Christchurch Call deserves recognition as a concrete, substantial achievement in several key areas. Firstly, as Peter Thompson argues, getting a coalition of state-actors and massive multinational corporations to publicly agree to basic principles on a subject as important and divisive as this one potentially matters more than the non-binding agreement in itself (Thompson 2019, 90) and is almost unprecedented. The Christchurch Call offers a platform that can be built upon and has laid the groundwork for future discussions.

Secondly, it achieved this through massive international public pressure in the wake of the Christchurch attacks that meant both governments and corporations felt vulnerable. The fact that it avoided typical pitfalls of responses to crises, where the affected industry announces initiatives internal to itself or where governments try to respond in isolation, is also notable. As a structural response to a very complex problem, the Call shows significant promise. Having a diverse plurality of global nation-states responding to international citizen-led pressure trying to solve a problem together is more likely[3] to reach robust solutions than alternative approaches centred on one country. Better still, the groups involved were powerful enough that technology companies who have historically refused to participate in these forms of discussions felt compelled to become involved. Perhaps most importantly, the Call is evidence that both governments and large corporations actually responded to sufficient levels of public pressure, suggesting that it is a method that can potentially achieve traction going forward.

Thirdly, the Call helped change the conversations unfolding regarding the role of social media in hate, harassment and terrorism. Jacinda Ardern, the Prime Minister of Aotearoa, has been publicly explicit in arguing that the Christchurch attack was 'shocking in its use of social media as a tool in the act of terror' (Ardern 2019a). Additionally, the Call makes specific mention of the algorithmic dimensions to online hate, requiring that Online Service Providers

> review the operation of algorithms and other processes that may drive users towards and/or amplify terrorist and violent extremist content to better understand possible intervention points and to implement changes where this occurs. This may include using algorithms and other processes to redirect users from such content or the promotion of credible, positive

> alternatives or counter-narratives. This may include building appropriate mechanisms for reporting, designed in a multi-stakeholder process and without compromising trade secrets or the effectiveness of service providers' practices through unnecessary disclosure. ('Christchurch Call' 2019)

Business models are also obliquely discussed in the section regarding responsibilities jointly held by both governments and Online Service Providers:

> Respect, and for Governments protect, human rights, including by avoiding directly or indirectly contributing to adverse human rights impacts through business activities and addressing such impacts where they occur. ('Christchurch Call' 2019)

Prior to the Christchurch Call, it was unthinkable that major technology companies could publicly acknowledge the possibility that their algorithms and fundamental business models might amplify terrorist content or not respect human rights. The Call explicitly makes them part of the conversation going forward (Cheng 2019a, 2019b).

However, there are also areas where the Christchurch Call significantly undermines its own stated aims.

The Flawed

The Christchurch Call conceptualises the dynamics of online harassment in narrow ways. The understanding that the Call presents of the 2019 attacks in Christchurch is that they were 'normal' terrorist violence unfolding in a country that does not normally experience them and that they were broadcast online. As such, the key focus is on preventing such broadcasts in future, since the understanding is that social media simply provided tools for magnifying the impact and reach of an act of terrorist violence that would have happened anyway. However, this understanding completely ignores the social dynamics which produced the Christchurch attack, of which the attack itself was merely the most visible dimension. The attacker wrote, 'Well lads, it's time to stop shitposting and time to make a real-life effort post' (Macklin 2019; Rowe 2019) as part of their manifesto: they explicitly address a broader online community in the attack and are courting social capital within it through the attack.

The Call shows no awareness of how Christy Dena's concepts of tiers[4] function within the context of harassment communities. As a result, it misses the dynamics by which the members of the primary, most active tier are encouraged and supported by the membership of less active tiers and compete with other members of the primary tier for visibility. It also misses that the fact the terror attack was live-streamed was itself designed to encourage people to join similar white-supremacist communities online. This recruitment would then expand the size of the tiers and thus the odds that others will be willing, able and motivated to enact further terror attacks. Instead, it seems to believe that the Christchurch terrorist attack came from an isolated group who happened to use social media as a tool.

Additionally, although the Call makes specific mention of the algorithmic dimensions to the amplification of extremist content online, the focus appears to be on stopping its distribution rather than, again, the dynamics which fuel it. There appears to be little focus on the ways that YouTube's algorithms are manipulated by, motivate and produce profit for networks of white-supremacists and other extremists (Lewis 2018). There is also no direct engagement with the fact social media companies currently profit directly from harassment.

The Christchurch Call's focus on terrorism also seems very literal: it focuses on what can easily be labelled as acts of political violence in the eyes of the public, but would seem completely blind to something like Kiwi Farms' organised attempts to drive marginalised people to suicide. It is also unclear as to whether Elliot Rodger's murderous misogynist rampage would be on the Call's radar either, despite the fact it and his incel manifesto have inspired other killings.

Beyond these conceptual issues, there are also practical concerns. The Christchurch Call is non-binding, but even so there is a complete lack of either a timeframe by which improvements might be assessed or information about who would be positioned to audit any claims for progress that are made.

The Actively Problematic

In addition to areas where the Call's specific policy frameworks undercut its stated aims, there are areas which either risk worsening the existing status quo or otherwise add new problems to the conversation.

One such example is that the summit began with the statement that the white-supremacist terrorist attack in Christchurch was 'unprecedented'

('Christchurch Call' 2019), but such a claim ignores a significant amount of history and context. The terrorism in Christchurch was not the first time that social media platforms have been implicated in terrorism: this is the first time that a terrorist attack in a 'western' country was broadcast via the internet, but Facebook has been a significant factor in the genocide of Rohingya Muslims in Myanmar, for example, as covered in the Frontline documentary *The Facebook Dilemma* (Jacoby 2018). Additionally, a study called 'Fanning the Flames of Hate: Social Media and Hate Crime' by Karsten Müller and Carlo Schwarz demonstrated a link between Facebook use and violence against refugees in Germany (Müller and Schwarz 2018; Taub and Fisher 2018). Social media has been connected to acts of terrorism and broader social abuses for a long time. Explicit acts of terrorism are merely the tip of an iceberg produced by a vastly more substantial broad base of problem content and communities—and this base is ignored by the current framing of the Christchurch Call. To claim that the attack is 'unprecedented' muddies important waters and endorses the idea that social media companies are blameless in both past events and the status quo. This is far from a neutral angle to be taking, particularly at a summit which is foundationally *about* preventing future terrorist attacks tied to social media.

Another area in which the Christchurch Call is in danger of simply adopting the terms preferred by the major social media platforms and big tech companies is around the positioning of how terrorism on social media is framed. Facebook COO Sheryl Sandberg made a public statement about the Christchurch Call and said,

> There was not one country represented in that room that had not been touched by terrorism and violent extremism. And the terrorist's goals are very clear: they aim to silence, they aim to stand against the values that we hold so dear, they aim to have people live in fear. And our goals are exactly the opposite—we want people to have voice, we want people to live with humanity and dignity. We want people to speak for tolerance and against hatred. (Cheng 2019c)

Sandberg positions Facebook as 'one of us' via a claim of shared values which stand against terrorism and the ideals of individual terrorists. Coincidentally, this positioning erases Facebook's extremely public controversies, many of which tie to its spread of intolerance and hatred and suggest that Facebook's pursuit of giving people 'voice' is a neutral good

rather than their pragmatic business model. As such, Sandberg's comments try to reframe the Christchurch Call as a rallying cry for the status quo and frame Facebook's business models as a brave defiance of a decontextualised terrorism.

Additionally, one of the risks presented by turning 'terrorist' into an empty-signifier is that recent history and the 'War on Terror' associates it disproportionately with Muslim communities. As a result, decontextualised use of the phrase to signify 'an enemy' runs the risk of reinforcing the racist and Islamophobic assumptions which motivated the attacks in Christchurch. Doing so would place the Christchurch Call on the side of the attacker rather than the communities which have been terrorised.

Since the Christchurch Call was announced, companies like Facebook have announced policy changes which would appear to contravene the agreements made in the Call—and these have gone unchallenged. For example, despite the driving force of the Call being to stop the spread of extremist content online, Facebook announced a revised 'newsworthiness' policy framing anything a politician says to be publishable, regardless of its content, in September 2019:

> Today, I announced that from now on we will treat speech from politicians as newsworthy content that should, as a general rule, be seen and heard. However, in keeping with the principle that we apply different standards to content for which we receive payment, this will not apply to ads—if someone chooses to post an ad on Facebook, they must still fall within our Community Standards and our advertising policies.
>
> When we make a determination as to newsworthiness, we evaluate the public interest value of the piece of speech against the risk of harm. When balancing these interests, we take a number of factors into consideration, including country-specific circumstances, like whether there is an election underway or the country is at war; the nature of the speech, including whether it relates to governance or politics; and the political structure of the country, including whether the country has a free press. In evaluating the risk of harm, we will consider the severity of the harm. Content that has the potential to incite violence, for example, may pose a safety risk that outweighs the public interest value. Each of these evaluations will be holistic and comprehensive in nature, and will account for international human rights standards. (Clegg 2019)

Despite the extensive reassurances of the consideration Facebook makes in deciding what is and is not 'newsworthy' under this policy, it is unclear

what the threshold is for action. For example, the Christchurch Call makes specific mention that Online Service Providers need to act to 'prevent the dissemination' of 'violent extremist content.' However, there have been no attempts to limit the dissemination of President Trump's statements on the platform which blame COVID-19 on China and Asian peoples, which have been connected to increases in xenophobic violence in an election year (Yam 2020).[5] If Facebook will not honour the Christchurch Call under the precise circumstances its own policy says that they will, it is unclear when they ever would.

The Risks and Rewards of Regulation

Given that the status quo of the social media landscape is both intolerable and dangerous, and that the Christchurch Call is problematic and underutilised, the question becomes what can be done to change things. Peter Thompson has presented an extensive, nuanced and clear discussion of the tensions surrounding regulation alongside the opportunities raised by the Christchurch Call in 'Beware of Geeks Bearing Gifts: Assessing the Regulatory Response to the Christchurch Call' (Thompson 2019), and I will explore some of his ideas here.

Thompson notes that historically, governments have allowed tech companies and social media platforms to largely regulate themselves due to their size, scope and complexity (Thompson 2019, 84). As a result, one of the difficulties of creating a more heavily regulated environment is that it is tempting for governments to delegate the specifics to companies who would then create regulation to suit existing market leaders to the detriment of new competition and diffusing their liabilities onto third parties (Thompson 2019, 85).

The Christchurch Call is highlighted as contributing by bringing diverse stakeholders across nation-states and major corporations to the table, and ensuring that discussions of regulation can no longer be preempted and shot down (Thompson 2019, 84, 99). As such, Thompson argues that it provides a useful foundation to build forward from, and he provides a specific framework for regulatory measures that would extend the Christchurch Call's strengths while mitigating its weaknesses.

Thompson notes that one of the reasons major technology companies and social media platforms were motivated to participate was in order to avoid having to respond to a patchwork of different regulatory measures in different countries (Thompson 2019, 92). This means that citizens

pushing for *more* such measures in different jurisdictions become a concrete, decentralised method of bringing further pressure to bear on them. Thompson advocates continued domestic regulatory response on these grounds, alongside the fact that any results from a multilateral forum are currently hypothetical and would need to be implemented domestically in any case (Thompson 2019, 92).

Another domestic response that will be valuable going forward is to levy taxes on the domestic turnover of global intermediaries. Thompson highlights that this has already happened in Britain and France in initiatives that are important because they 'reclaim online commercial turnover as domestic economic activity' (Thompson 2019, 86–87). Australia has moved to force Facebook and Google to share advertising revenue with local news media after attempts to collaboratively produce a voluntary code failed in 2019 (Jose and Packham 2020; Thompson 2019, 88). I argue that one of the problems in dealing with major technology companies from a domestic perspective is that there can be no political will to regulate companies if there is no political will to tax them. Taxing companies helps correct fundamentally extractive business models that treat the activity of a given nations' citizens as valuable but returns none of that value to their nations' economy. In addition, it represents exactly the kind of piecemeal regulatory response that major technology companies and social media platforms would prefer to avoid through engaging in more unified responses, at the same time as being valuable to the individual countries setting the levy.

Thompson's work provides a roadmap for wider regulatory responses than taxation that different territories can follow as well. These include responding to the concentration of content-discovery and e-commerce by 'redesignating digital intermediaries as public utilities with civic obligations beyond private shareholders,' and independent regulator access to algorithms (Thompson 2019, 97–98). Effectively, this is an area where the citizens of the world have been presented as powerless for decades and instead have some options for local political activism. It is possible to build a sufficient diversity of sticks to wield against big players in social media and technology that the carrot of greater simplicity in cooperating with multilateral regulation becomes more attractive as a result.

Tackling Invisible Problems

Alongside applying regulatory pressure to politicians in our own domestic contexts, we can work against the fact that the ideologies which drive harassment campaigns and crowdsourced terrorism are *mainstream*. As such, they are invisible problems that are easy to overlook, precisely because they are part of the background-radiation of everyday life.

We can look to the media landscape of Aotearoa-New Zealand for a very specific example of this kind of dynamic. On 28 August 2018, columnist and broadcaster Mike Hosking wrote a column for the *New Zealand Herald* that argued Chelsea Manning should be barred from entering Aotearoa because it was reprehensible to allow a criminal to profit from criminal behaviour (Hosking 2018). That column draws an explicit comparison between Chelsea Manning and the white-supremacist duo Lauren Southern and Stefan Molyneux, who were banned from speaking publicly after an outcry during their visit to Aotearoa (Hatton 2018). The comparison is brief, serving to set up a binary whereby Manning is only notable due to criminality, while Southern/Molyneux are not criminals and thus should be allowed to speak.

> If it wasn't for the stealing and leaking of classified paperwork that ran the risk of undermining American security, you would never have heard of her. Far less be in a position to consider buying tickets and lining her, and her promoters', pockets.
>
> Which brings us to free speech—the same free speech we were angsting about a few weeks ago when those Canadian right wingers Lauren Southern and Stefan Molyneux were here.
>
> I support free speech, I would have let those two in, not because I have any particular interest in what they have to say, but because they are free to say it. And if we are to choke off all discourse every time it might look like we don't like what is being said, we are on a very slippery slope.
>
> Which is what made Phil Goff's moves so egregious, and every other hand-wringer that lined up behind him. (Hosking 2018)

The clear implication is that Southern and Molyneux may have controversial views, but they are not *criminals*. However, Southern was arrested in 2017 for being part of a group that blocked rescue boats in the Mediterranean that were seeking to save immigrants drowning at sea—an event that she live-streamed (Townsend 2017; Warren 2017). She and Molyneux are central parts of the alt-right/white-supremacist influencer

network identified by Rebecca Lewis (Lewis 2018) and promote the 'Great Replacement Theory' that was cited by the Christchurch terrorist as justification for the attacks.

For clarity, I do not argue that Hosking's column contributed to the Christchurch terror attacks, but it highlights the invisibility, insulation and normalisation of white-supremacy and white-nationalism within mainstream culture. Importantly, the issue is not that these problems with Southern and Molyneux were known but ignored. I think it is more likely that everyone involved in producing the column was unaware because Southern and Molyneux blend in so that there was no reason to look deeper, despite the fact their views qualify as hate speech according to experts within Aotearoa (Hatton 2018; 'Kōrero Whakamauāhara: Hate Speech—An Overview of the Current Legal Framework' 2019, 4, 7). It is possible to be detained for attempting to disrupt humanitarian search and rescue efforts—an initiative which, if successful, would directly lead to the deaths of people who might otherwise be saved—and not be a 'criminal.' It is possible to publicly argue that indigenous people are the 'lowest rung of civilization' (NZ Media Council 2018) and that non-white cultures and peoples are an existential threat to society that must be violently resisted, and it is just something 'we don't like being said' (Hosking 2018). Selling tickets to public events where these ideas will be promoted at NZ$99 per ticket is not profiting from criminal enterprise, whereas whistleblowing against government overreach is a bridge too far. It is possible to publish a column in a major national newspaper that uses people promoting hate speech for money as a minor side note to say 'Chelsea Manning isn't like them: *she's* actually a problem,' and there is no consumer outcry or backlash from advertisers because it is *normal*. It does not stand out.

Aotearoa-New Zealand makes for a specific example of these dynamics because they are invisible even when violently contrasted with reality. The country was outraged by the Christchurch attacks, insisting that 'this is not us'—a claim vigorously challenged by Māori alongside Asian and Muslim communities of Aotearoa (Bashir 2019; Han 2019; McLachlan 2019). As with many Commonwealth countries, Aotearoa-New Zealand was founded as a white-supremacist, colonialist project, and that foundation extends to the present day in most areas of life. It is only recently that te reo Māori has begun to be taught in schools, despite being listed as one of the 'official' languages of Aotearoa, and mainstream media representations position Māori as dangerous 'others' who are a threat to 'the nation' (Abel 2013).[6] One of our more influential political parties owes most of its

identity and branding to being anti-immigration, with its members of parliament regularly voicing Islamophobic or anti-Asian statements. Aotearoa-New Zealand is far from unique in displaying rich veins of normalised, mainstream racism, because all of the problems discussed here are global.

Hate, discrimination and harassment are fractal, in that the same patterns play out again and again at different scales and in different contexts. In May 2020, NBC revealed that Google had rolled-back diversity initiatives because of pressure from those who argued that employing and supporting people who were not straight, white, cisgender men was 'anti-conservative' (Glaser 2020). Such a claim explicitly admits that modern conservatism is a sexist, racist, patriarchal social project and one which lines up almost perfectly with the ideologies displayed by online harassment communities—partly because these same people could easily be part of both. As the title of Adrienne Shaw's rich and detailed article says, 'The Internet Is Full of Jerks Because the World Is Full of Jerks' (Shaw 2014), and once again, the question becomes what to do about it. As this book has repeatedly argued, there is no segregation between online spaces and the 'real world,' which means our responses to cultures of harassment cannot be entirely focused in one area either. Saziah Bashir's recommendations for tackling cultures of racism and inequity after the Christchurch attacks are broadly applicable and seem like an excellent place to start:

> As a society, we can choose which voices we empower and nurture at critical junctures that can shift the conversation, and we must be diligent and deliberate in exercising that choice now.
>
> (…) speak out. Challenge destructive narratives in the media, from your community leaders and politicians, and the people in your life.
>
> Call up and debate with those talkback radio hosts, argue with your racist uncle at that family barbecue, tell those problematic old high school friends exactly why you're blocking and deleting them before you do it, disparage your friends from laughing at that racist joke or using that unacceptable word, ask your employer to account for the efficacy and fairness of their recruitment policies and their commitment to diversity, attend the rallies and sign the petitions.
>
> If someone is not good enough to be your leader, then get them sacked (more effective than egging, though visually less striking).
>
> If that sounds like a lot of work, it is. If you want to be an ally, do the work. People in marginalised communities have been doing it all along for our own people and often everyone else. If you are born into an identity

> whose intersections suffer little or no disadvantage or discrimination, do more for others with your privilege. (Bashir 2019)

Alongside tackling the background-radiation of ubiquitous social sexism, racism, ableism, homophobia and transphobia, we can push for structural changes in governments and corporations. Connecting back to Golding and Van Deventer's point back from Chap. 1:

> Maybe it's most useful not to look at a chronology of abuse to work out what's encouraging such behaviour, but rather to look at the systems surrounding this abuse. What stays the same over the years? The targets change. The harassers change. But the systems that harbour the behaviour of the harassers haven't changed enough. Unless we get real systemic change through strong and assertive leadership in tech companies, not much is going to improve for women and minorities online. (Golding and Van Deventer 2016, 101)

This book has explored the ways that the pragmatic structure of social media platforms and online spaces is as relevant to harassment as the social contexts they operate within, and the two are innately linked. Much of the consistent blindness displayed by the technology industries regarding harassment and the way their design decisions can be weaponised by bad actors would be resolved by a greater diversity of people involved in the decision-making.

Pushing for substantive, meaningful changes at these levels will be challenging, particularly in the international political climate of 2020 and the years to come. The rise of interest in unionisation in technology and creative industries is a promising opportunity, although it will be as necessary to hold those unions themselves accountable on issues of diversity and representation as it is for the corporations those unions engage with.

The status quo is untenable and harms people.
Nothing about the status quo is a neutral inevitability: it can be changed.
However, it is not going to change by itself, people will have to make that happen.
Hopefully, this book offers some insights into what that might look like.

Notes

1. The initial nation-states who signed the Christchurch Call were Aotearoa-New Zealand, Australia, Canada, the European Commission, France, Germany, Indonesia, India, Ireland, Italy, Japan, Jordan, the Netherlands, Norway, Senegal, Spain, Sweden and the United Kingdom. By September 2019, this had expanded to a total of 47 countries. The United States cited support for the summit but claimed to be constrained by the First Amendment—a claim already challenged by Danielle Keats Citron, who argues that preventing hate speech and online harm is thoroughly consistent with the First Amendment (Citron 2014, 190–225).
2. The corporate signatories were Amazon, Daily Motion (owned by Vivendi), Facebook, Google, Microsoft, Qwant (a French search engine), Twitter and YouTube (a Google subsidiary).
3. Or at least, *less unlikely*.
4. See Chap. 3.
5. Twitter is not currently a signatory of the non-binding Christchurch Call, but another example of this kind of dynamic is that Twitter has refused to apply the same techniques that remove pro-ISIS accounts from the network to white-supremacists because doing so would also remove Republicans (Cox and Koebler 2019). Likewise, there is the example where a Twitter account resharing content from Donald Trump with no alterations was suspended after operating for just 68 hours on the grounds it was 'glorifying violence,' while Trump's account itself was defended as 'public interest' (Yeo 2020). And again, we can interrogate why Twitter removes and filters algorithmically identified white-supremacist and neo-Nazi accounts in France and Germany where required by law, but does not apply the same filter to the network globally.
6. The New Zealand First party and all its members were removed from parliament as a result of a national election in 2020. However, they still exist as a political organization and may contest future elections.

References

Abel, Sue. 2013. Māori, Media and Politics. In *Politics and the Media*, 257–271. Auckland: Pearson.

Ardern, Jacinda. 2019a. NZ and France Seek to End Use of Social Media for Acts of Terrorism. *The Beehive*, 24 April 2019. http://www.beehive.govt.nz/release/nz-and-france-seek-end-use-social-media-acts-terrorism.

———. 2019b. Christchurch Call to Eliminate Terrorist and Violent Extremist Online Content Adopted. *The Beehive*, 16 May 2019. http://www.beehive.

govt.nz/release/christchurch-call-eliminate-terrorist-and-violent-extremist-online-content-adopted.

Bashir, Saziah. 2019. Four Things You Should Do Following the Christchurch Terror Attacks. *RNZ* (19 March 2019) https://www.rnz.co.nz/news/on-the-inside/385064/saziah-bashir-four-things-you-should-do-following-the-christchurch-terror-attacks.

Cheng, Derek. 2019a. Christchurch Call to Action: Govts, Tech Companies Agree to Tackle Violent Online Content on Social Media. *NZ Herald*, 15 May 2019, sec. New Zealand, Northern Advocate. https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=12231337.

———. 2019b. Profit-Chasing Leads YouTube Users to Radicalisation and Conspiracy Theories, Say Tech Experts. *NZ Herald*, 15 May 2019, sec. Business. https://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=12230981.

———. 2019c. Christchurch Call Update: Social Media Giants Join Forces to Fight Extremism—NZ Herald. *NZ Herald*, 24 September 2019. https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=12269942.

'Christchurch Call'. 2019. The Christchurch Call. 15 May 2019. https://www.christchurchcall.com/call.html.

Citron, Danielle Keats. 2014. *Hate Crimes in Cyberspace*. Cambridge, MA: Harvard University Press.

Clegg, Nick. 2019. Facebook, Elections and Political Speech. *About Facebook* (blog). 24 September 2019. https://about.fb.com/news/2019/09/elections-and-political-speech/.

Cox, Joseph, and Jason Koebler. 2019. Twitter Won't Treat White Supremacy Like ISIS Because It'd Have to Ban Some GOP Politicians Too. *Vice* (blog). 25 April 2019. https://www.vice.com/en_us/article/a3xgq5/why-wont-twitter-treat-white-supremacy-like-isis-because-it-would-mean-banning-some-republican-politicians-too.

Diakopoulos, Nicholas. 2015. Algorithmic Accountability. *Digital Journalism* 3 (3): 398–415. https://doi.org/10.1080/21670811.2014.976411.

Glaser, April. 2020. Current and Ex-Employees Allege Google Drastically Rolled Back Diversity Programs. *NBC News*, 13 May 2020. https://www.nbcnews.com/news/us-news/current-ex-employees-allege-google-drastically-rolled-back-diversity-inclusion-n1206181.

Golding, Dan, and Leena Van Deventer. 2016. *Game Changers: From Minecraft to Misogyny, the Fight for the Future of Videogames*. South Melbourne: VIC Affirm Press.

Han, William. 2019. Is New Zealand Welcoming of All Races? Look at What Its Politicians Have Said. *South China Morning Post*, 17 March 2019. https://www.scmp.com/week-asia/opinion/article/3002037/christchurch-shooting-racism-new-zealand-isnt-new-and-era-donald.

Hatton, Emma. 2018. Far-Right Pair Banned from Speaking at Auckland Council Venues—Phil Goff. *RNZ*. 6 July 2018. https://www.rnz.co.nz/news/national/361220/far-right-pair-banned-from-speaking-at-auckland-council-venues-phil-goff.

Hosking, Mike. 2018. Mike Hosking: Chelsea Manning Is a Crook, Keep Her out of NZ. *NZ Herald*, 28 August 2018, sec. New Zealand. https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=12115380.

Jacoby, James. 2018. The Facebook Dilemma. *Frontline*. *PBS*. https://www.pbs.org/wgbh/frontline/film/facebook-dilemma/.

Jose, Renju, and Colin Packham. 2020. Australia to Force Google, Facebook to Pay Domestic Media to Use Content. *Reuters*, 20 April 2020. https://www.reuters.com/article/us-australia-media-regulator-idUSKBN222066.

'Kōrero Whakamauāhara: Hate Speech—An Overview of the Current Legal Framework'. 2019. Aotearoa-New Zealand: Human Rights Commission (Te Kāhui Tika Tangata).

Lewis, Rebecca. 2018. Alternative Influence: Broadcasting the Reactionary Right on Youtube. *Data & Society*. https://datasociety.net/output/alternative-influence/.

Macklin, Graham. 2019. The Christchurch Attacks: Livestream Terror in the Viral Video Age. *Combating Terrorism Center at West Point* (blog). 18 July 2019. https://ctc.usma.edu/christchurch-attacks-livestream-terror-viral-video-age/.

McLachlan, Leigh-Marama. 2019. Christchurch Mosque Attacks: Māori Leaders Say Acts of Terror Nothing New in New Zealand. *RNZ*. 21 March 2019. https://www.rnz.co.nz/news/national/385226/christchurch-mosque-attacks-maori-leaders-say-acts-of-terror-nothing-new-in-new-zealand.

Müller, Karsten, and Carlo Schwarz. 2018. Fanning the Flames of Hate: Social Media and Hate Crime. *SSRN*, November. https://doi.org/10.2139/ssrn.3082972.

NZ Media Council. 2018. Robin Grieve Against New Zealand Herald. *New Zealand Media Council* (September 2018) https://www.mediacouncil.org.nz/rulings/robin-grieve-against-new-zealand-herald-2.

Rowe, Don. 2019. The Online Cesspits Where Hate Found a Home. *The Spinoff* (blog). 19 March 2019. https://thespinoff.co.nz/media/19-03-2019/the-online-cesspits-where-hate-found-a-home/.

Shaw, Adrienne. 2014. The Internet Is Full of Jerks, Because the World Is Full of Jerks: What Feminist Theory Teaches Us About the Internet. *Communication and Critical/Cultural Studies* 11 (3): 273–277. https://doi.org/10.1080/14791420.2014.926245.

Taub, Amanda, and Max Fisher. 2018. Facebook Fueled Anti-Refugee Attacks in Germany, New Research Suggests. *The New York Times*, 21 August 2018, sec. World. https://www.nytimes.com/2018/08/21/world/europe/facebook-refugee-attacks-germany.html.

Thompson, Peter A. 2019. Beware of Geeks Bearing Gifts: Assessing the Regulatory Response to the Christchurch Call. *The Political Economy of Communication; Vol 7, No 1 (2019)*, August. http://www.polecom.org/index.php/polecom/article/view/105/314.

Townsend, Mark. 2017. Far Right Raises £50,000 to Target Boats on Refugee Rescue Missions in Med. *The Observer*, 3 June 2017, sec. World News. https://www.theguardian.com/world/2017/jun/03/far-right-raises-50000-target-refugee-rescue-boats-med.

Warren, Rossalyn. 2017. Europe's Far-Right Pirates of the Mediterranean Are Targeting Refugee Rescue Missions. *Washington Post* (28 July 2017) https://www.washingtonpost.com/news/global-opinions/wp/2017/07/28/europes-far-right-pirates-of-the-mediterranean-are-targeting-refugee-rescue-missions/.

Yam, Kimmy. 2020. Trump Doubles down That He's Not Fueling Racism, but Experts Say He Is. *NBC News*, 18 March 2020. https://www.nbcnews.com/news/asian-america/trump-doubles-down-he-s-not-fueling-racism-experts-say-n1163341.

Yeo, Amanda. 2020. One Twitter Account Is Reposting Everything Trump Tweets. It Was Suspended within 3 Days. *Mashable*, 3 June 2020. https://mashable.com/article/twitter-donald-trump-suspend-tweets-policy-violence/.

Index[1]

[1] Note: Page numbers followed by 'n' refer to notes.

K. Veale, *Gaming the Dynamics of Online Harassment*,
https://doi.org/10.1007/978-3-030-60410-3