

Northumbria Research Link

Citation: Are, Carolina (2023) An autoethnography of automated powerlessness: lacking platform affordances in Instagram and TikTok account deletions. *Media, Culture & Society*, 45 (4). pp. 822-840. ISSN 0163-4437

Published by: SAGE

URL: <https://doi.org/10.1177/01634437221140531>
<<https://doi.org/10.1177/01634437221140531>>

This version was downloaded from Northumbria Research Link:
<https://nrl.northumbria.ac.uk/id/eprint/50633/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

An autoethnography of automated powerlessness: lacking platform affordances in Instagram and TikTok account deletions

Media, Culture & Society

2023, Vol. 45(4) 822–840

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/01634437221140531

journals.sagepub.com/home/mcs**Carolina Are** 

Northumbria University, UK

Abstract

Situated within the field of platform governance studies, this paper shares findings from an ‘autoethnography of automated powerlessness’, drawing from the researcher’s disempowering experience of being a heavily moderated social media user. Using theoretical frameworks blending affordances and World Risk Society theories, this paper contextualises my experiences of moderation of my pole dance instructor, activist and blogger account @bloggeronpole from February to October 2021 within social media’s broader de-platforming of nudity and sexuality, finding fallacies within platforms’ own affordances, which lack mechanisms to aid or rehabilitate de-platformed accounts. With little to no information from platforms about the details of their moderation, qualitative, ethnographic and autoethnographic explorations of their governance are all users currently have to fight and understand their puritan, patriarchal censorship of nudity and sexuality, which are often conflated with risk. This study concludes with recommendations for different options for better, more equal and community focused moderation.

Keywords

algorithm bias, automated content moderation, deplatforming, Instagram, platform governance, social media moderation, social media nudity, TikTok

Corresponding author:

Carolina Are, Centre for Digital Citizens, Northumbria University, Northumberland Road, Newcastle upon Tyne NE1 8ST, UK.

Email: carolina.are@northumbria.ac.uk

Introduction

This paper is situated within the platform governance studies and online moderation space. For Tiidenberg, platform governance can both be an area of research and activism, as it ‘comprises questions pertaining to the implications and impact of platform features, functions and rules’ as well as ‘the international regulatory dynamics that currently delineate the freedoms, responsibilities and liabilities of platform companies’ (Tiidenberg, 2021: 2).

The moderation – for example, the deletion or censorship – of online content is a key aspect of platform governance. Without online moderation, Diaz and Hecht-Felella (2021) write, platforms would be unusable. And while platforms initially rejected claims that their power was akin to those of publishers, preferring to define themselves as utilities or tools (Gillespie, 2010; Zuckerberg, 2018), they nonetheless make publisher-like decisions over what types of content is allowed and visible in their spaces through the affordances – or the possibilities objects, and therefore technologies, offer for action (Graves, 2007; Norman, 1988) – of their governance infrastructure. Said decisions, this paper will argue, are an online translation of World Risk Society theory, based on the idea that institutions and businesses ineffectively attempt to reduce risks for their citizens or customers by restricting civil liberties (Beck, 1992, 2006; Giddens, 1998; Hudson, 2003).

Platforms’ content moderation has already been identified as a vehicle for harm when preventing users to communicate and organise in critical situations. For example, it has been found that Meta’s 2019 banning of organisations in conflict with the Myanmar military affected these organisations’ ability to seek help from international stakeholders (Sablosky, 2021). Even outside conflict scenarios, research has found repeated examples of online moderation targeting marginalised users, and disproportionately focusing on nudity and sexuality instead of on violence, particularly after the approval of FOSTA/SESTA in the United States (Are, 2020c; Diaz and Hecht-Felella, 2021; Haimson et al., 2021; Paasonen et al., 2019; Tiidenberg and van der Nagel, 2020; etc.). After the US Congress approved FOSTA/SESTA, a 2018 exception to Section 230 of the Telecommunications Act which has kept online services immune from liability for the actions of their users, social media moderation of nudity changed (Are, 2021c; Are and Paasonen, 2021; Paasonen et al., 2019; Tiidenberg and van der Nagel, 2020). To avoid being seen to be facilitating sex trafficking, social media platforms began to increasingly censor nudity and sex, affecting first sex workers, then pole dancers, athletes, sex educators and anyone displaying their body online (Are, 2021c; Are and Paasonen, 2021; Paasonen et al., 2019; Tiidenberg and van der Nagel, 2020). Given the importance of social media as work, promotional and civic spaces (Are, 2020a, 2020b, 2020c; Banet-Weiser, 2018), this heightened moderation of nudity has been affecting the livelihoods, lives and visibility of a set of users, who find themselves powerless in trying to understand or appeal these decisions, and in attempting to re-gain control over their profiles, content and social media history (Are and Paasonen, 2021; Blunt and Wolf, 2020; Duffy and Meisner, 2022).

To highlight the challenges users face when dealing with automated moderation, this paper explores the disempowering aftermath of losing one’s profile from an autoethnographic perspective, through the researcher’s own experience of Instagram and TikTok

account deletions. With little to no information from social media companies about the details of their moderation, qualitative, ethnographic and autoethnographic explorations of platform governance are all users currently have to fight and understand their puritan censorship of nudity and sexuality. This paper contextualises this experience through affordances (Graves, 2007; Norman, 1988) and World Risk Society (Beck, 2006; Giddens, 1998; Hudson, 2003) theories. It finds platforms' affordances to aid de-platformed users lacking and concludes by recommending different options for better, more equal and community focused moderation.

Social media moderation research has so far explored platforms' 'shadowbanning' of content and profiles, hiding users from apps' main feeds without their knowledge and de facto limiting their visibility and reach (Are, 2021c; Are and Paasonen, 2021; Cotter, 2021; Savolainen, 2022), as well as practices such as 'flagging' (Crawford and Gillespie, 2016; Fiore-Silfvast, 2012; Peterson, 2013). Further studies focused on the 'assembling' of social media moderation and content recommendation processes to perpetuate sexism (Gerrard and Thornham, 2020), on the pitfalls and biases of automated moderation (Binns, 2019; Gillespie, 2018; Kaye, 2019; Paasonen et al., 2019; etc.) and on platform governance of creators at the margins (Duffy and Meisner, 2022). Research also examined the censorship of sex or sexual content (Tiidenberg and van der Nagel, 2020) and its offline consequences among sex worker communities (Blunt and Wolf, 2020). However, while social media platforms' overall lack of transparency, communications and accountability have been constantly mentioned throughout online moderation research, studies have yet to examine the challenges platforms affordances pose to users once they are de-platformed and directly faced with said lack of transparency, communications and accountability.

The lacking affordances of content moderation

This paper conceptualises the effects of platform governance from a user perspective, defining it as 'automated powerlessness' arising from platforms' own affordances. Researching the effects of social media's affordances is crucial not just to understand what types of action platform infrastructure enables: indeed, in shaping the norms of behaviour on their sites, social media make political decisions about what is acceptable, de facto ruling over ideas and beliefs in an increasingly essential civic space where we work and express ourselves (Are, 2020a, 2020b, 2020c; Crawford and Gillespie, 2016; Gillespie, 2010).

While initially social media movements such as the Arab Spring and #OccupyWallStreet positioned social networks as an opportunity to bypass gate-keepers and give a voice to the voiceless (Sloan and Quan-Haase, 2017; Vivienne, 2016), increasing concerns about online harms such as misinformation, conspiracy theories, hate speech and online abuse dominated the discourse about these platforms from the second half of the 2010s going forwards (Bartlett, 2018; Jane, 2014). These harmful behaviours are not strictly generated by platforms, but they can be exacerbated by their affordances, which include 'visible network relationships, quantified social endorsement (e.g. 'likes' and follows), and algorithmic feeds designed to maximize social engagement' (Schoenebeck and Blackwell, 2021: 2).

Still, research on and discussions surrounding platform affordances often focus on what said affordances allow users *to do*, instead of what they *prevent them from doing*, even though prevention – and therefore the inhibition of platforms’ own affordances – is at the centre of platform governance. Indeed, given the scale at which social media platforms have grown and operate, their governance and harm prevention infrastructure have had to rapidly develop from volunteer moderators towards automated moderation (Diaz and Hecht-Felella, 2021; Schoenebeck and Blackwell, 2021). Automated content moderation consists in the use of ‘machine learning and natural language processing to develop computational models that systematically evaluate large quantities of data’ (Schoenebeck and Blackwell, 2021: 6). Automated moderation is governed by rules known as ‘community standards’, through which platforms explain the kinds of content they allow on their apps, typically containing ‘rules against hate speech, terrorist or extremist content, nudity, and harassment’ (Diaz and Hecht-Felella, 2021: 4). Platforms’ community guidelines differ in wording but mirror one another (Diaz and Hecht-Felella, 2021).

However, for Schoenebeck and Blackwell, these guidelines are ‘not made transparent to users, both in process and outcome’, and they are enforced through the outsourcing of content reviewing to ‘third-party contractors who earn relatively low wages for work that is both physically and emotionally taxing’ (Schoenebeck and Blackwell, 2021: 2). Therefore, in practice, those enforcing most platforms’ community guidelines have so far struggled to distinguish between troubling content and content in the public interest, or content covered by freedom of expression rights (Gillespie, 2010; Kaye, 2019).

The many cases of mis-moderation related to the reliance on this un-trained and increasingly precarious workforce mixed with automated governance has resulted in calls for bigger investments in human moderators for fairer content moderation (Binns, 2019; Suri and Gray, 2019). The governance systems enforcing social media’s community standards have been deemed so opaque, unclear and inconsistent that they encourage censorship, hate speech and disinformation (Kaye, 2019; Paasonen et al., 2019). Therefore, as Schoenebeck and Blackwell wrote: ‘accurate and reliable detection is challenging at best, even in far less complex applications than the detection of nuanced behaviors like online harassment and hate speech’ (Schoenebeck and Blackwell, 2021: 6).

In this sense, platforms’ own *moderation* affordances, which rely on one-size-fits-all community guidelines largely enforced by automated processes more than by contextually trained human moderators, seem to be struggling to cope with the results of the *content creation* and *sharing* affordances their infrastructure has enabled.

The affordances of nudity governance on Instagram and TikTok

This paper focuses on platform governance of nudity and sexuality on Instagram and TikTok. Known to host celebrities and influencers, ‘highly visible tastemakers who professionally publish content on social media platforms’ and who can be seen as having Instagram and TikTok as their main ‘workplace’ (Are and Paasonen, 2021; Arriagada and Bishop, 2021: 1), the Meta-owned app and Chinese platform are currently two of the fastest growing social networks (Stefanello, 2021; Stokel-Walker, 2021). While these

platforms have been known to remove single pieces of content or to ‘shadowban’ profiles or posts – or to hide them without removing them or notifying users of said actions (Are, 2021a, 2021b; , 2021c; , 2021d; , 2021e; Savolainen, 2022) – this study’s main focus is on automated account deletions for nudity or sexual activity, enabled by a series of Instagram and TikTok affordances showcased in this section.

Both Instagram and TikTok disable or delete accounts that do not follow their community guidelines. Instagram (n.d.a, n.d.b) write that: ‘Accounts that don’t follow our Community Guidelines or Terms of Use may be disabled without warning’ and that they may ‘permanently remove an account that repeatedly violates the Community Guidelines or Terms of Use’. TikTok (n.d.), who claim that their algorithms are designed with ‘trust and safety in mind’, promise to ‘suspend or ban accounts and/or devices that are involved in severe or repeated violations’ of community guidelines. While reading these platforms’ rules, one would be forgiven for automatically assuming that nudity and sexuality are synonymous with risk, and that they represent an absence of safety.

Instagram’s community guidelines in relation to nudity are situated underneath a heading asking users to: ‘Post photos and videos that are appropriate for a diverse audience’ (Instagram, n.d.a, n.d.b). Nudity is, for them, not appropriate: ‘[A]lways follow the law’, say Instagram’s community guidelines, adding: ‘Respect everyone on Instagram, don’t spam people or post nudity’.

In an extraordinary juxtaposition of completely different activities, the rule stating that: ‘Instagram is not a place to support or praise terrorism, organised crime or hate groups’, is followed by the platform banning the offering and selling of sexual services, although different forms of sex work are legal in many countries (Paasonen et al., 2019; Tiidenberg, 2021; etc.). Indeed, ‘offering sexual services’ is not allowed on Instagram, together with ‘buying or selling firearms, alcohol and tobacco products between private individuals, and buying or selling non-medical or pharmaceutical drugs are also not allowed’ (Instagram, n.d.a, n.d.b). TikTok’s community guidelines with regards to nudity, too, prohibit posting anything that is ‘overtly revealing of breasts, genitals, anus, or buttocks, or behaviors that mimic, imply, or display sex acts’, because they are ‘mindful of content that may be offensive or culturally inappropriate in certain regions or may not be suitable for users of all ages’ (TikTok, n.d.).

Although community guidelines outline the content that platforms ban, Instagram and TikTok have repeatedly been found to de-platform content within their rules (Are, 2021a, 2021b, 2021c, 2021d, 2021e; Duffy and Meisner, 2022; etc.). Yet, the opacity of social media infrastructure means users are not always privy to what exactly triggered their account’s deletion (Schoenebeck and Blackwell, 2021): although both Instagram and TikTok do show users when their accounts rack up multiple violations and may be at risk of deletion (Kastrenakes, 2019), users do not have access to specific information, for example, whether deletions were caused by a single post, a succession of posts or a series of reports by other users. Platforms have been known to automatically disable profiles without warnings or explanations, often restoring them when faced with the PR damage ensuing from high-profile deletions (Are, 2021d; Stefanello, 2021). This is particularly worrying in the case of nudity and sexuality, which have in the past been the target of malicious reporting, or ‘flagging’ actions (Clark-Flory, 2019).

‘Flagging’ is a specific platform affordance consisting in ‘a mechanism for reporting offensive content to a social media platform’ (Crawford and Gillespie, 2016: 411). An important tool to single out online abuse, flagging has been exploited maliciously, in what Fiore-Silfvast (2012) has called ‘user-generated warfare (UGW)’, where targeted efforts by multiple accounts join forces to remove profiles whose content they disagree with. UGW is a serious danger for accounts posting nudity and sexuality, which are already under threat of de-platforming by social media community guidelines (Are, 2021a, 2021b, 2021c, 2021d, 2021e).

Flags are, for Crawford and Gillespie (2016: 419), a convenient mechanism for platforms to avoid government oversight, meaning that by enabling flagging they make ‘ad hoc and often self-interested assessments’ of whether content should, in fact, be removed – something very likely in the case of posts containing sex work, nudity and sexuality post-FOSTA/SESTA (Blunt and Wolf, 2020), particularly when the users who share them are women who are not famous. Indeed, while nudity can be lucrative for platforms, it is celebrated only when posted by a specific, verified or ‘safe’ type of user: community guidelines restrict female nipples and female nudity, but they allow male bodies and celebrities broader self-expression (Are and Paasonen, 2021; Paasonen et al., 2019). And while platform censorship has been found to target male users, specifically from BIPOC and LGBTQIA+ communities (Haimson et al., 2021), heterosexual, cis-gender male nudity has so far largely gone undetected by platforms, so much so that users have been able to post or direct message their penises without being algorithmically censored (Center for Countering Digital Hate, 2022; Peake, 2022). Conversely, creators who publish nuanced content featuring nudity and sex, toeing the line of platforms’ notions of acceptability, may be excluded from the visibility and work opportunities that influencer status provides, and be excluded from these social media spaces altogether:

In expensive interpretations of FOSTA/SESTA, sex workers are seen to sell their bodies and to advertise their services merely by virtue of their social media presence. This presence then becomes seen as problematic as such – even if the services that they promote, such as stripping, are legal in many countries, the USA included’ (Are and Paasonen, 2021: 8).

Although based in different countries – with Instagram originating in the US and TikTok in China – these apps’ approach to the moderation of nudity and sexuality reflects a typically North American mentality set by FOSTA/SESTA, which is having an impact on the visibility of nudity and sexuality on apps used around the world (Kaye, 2019; Paasonen et al., 2019; Tiidenberg and van der Nagel, 2020). Instagram and TikTok’s community guidelines therefore rely on old-fashioned and non-inclusive depictions of bodies: leaked reports showed Meta based their Community Standards on advertising guidelines by Victoria’s Secret (Salty, 2019) or tried to govern the limits of different modes of breast cupping or grabbing (Gilbert, 2020) using standards more akin to sexist advertising than to the progressive sexual practices showcased by platforms’ own users (Paasonen et al., 2019). Because of this approach to content governance sex and sexuality as risky and harmful, social media’s content moderation can be defined as ‘puritan’, characterised by ‘wariness, unease, and distaste towards sexual

desires and acts deemed unclean and involving both the risk of punishment and the imperative for control', so much so that sexuality must be feared, governed and avoided (Paasonen et al., 2019: 169). This, for Paasonen et al. (2019), undermines the centrality of sex in people's lives.

The politics of de-platforming sex in the World Risk Society

Since Instagram and TikTok's community guidelines position the need for 'safety' as the main driver of their governance of both online abuse and nudity, World Risk Society theory (Beck, 1992, 2006; Giddens, 1998; Hudson, 2003) is appropriate to understand social media account deletions.

Based on the idea that institutions and businesses ineffectively attempt to reduce risks for their citizens or customers by restricting civil liberties, World Risk Society theory sees corporations' attempts to avoid undesirable events by arbitrarily identifying risks to prevent, increasing the marginalisation of society's 'others' in doing so (Beck, 1992, 2006; Giddens, 1998; Hudson, 2003).

Risks are undesired, threatening events (Hudson, 2003) that become apparent through what Beck calls 'techniques of visualization', – for example, the mass media (Beck, 2006: 332). Modern society has become so preoccupied with risks that, for Beck, it has become a risk society 'debating, preventing and managing risks that it itself has produced' (Beck, 2006). This emphasis on risk brings society's main institutions to scramble to 'attempt to anticipate what cannot be anticipated' (Beck, 2006: 329), anticipating the wrong risks without preventing disasters arising from those that were not foreseen. This approach has mostly benefited private companies such as insurance firms, who identify risks to prevent and undesirables to exclude from society in the name of safety (Hudson, 2003). In a similar vein, social media companies – corporate entities in charge of a civic space that is being used for expression and debate (Are, 2020c; Bartlett, 2018; Kaye, 2019) – have responded to risks by introducing and changing their community guidelines, governance mechanisms enforced through platform affordances such as automated moderation (Are, 2020c; Bartlett, 2018; Kaye, 2019) that has so far removed content and accounts by some of society's most vulnerable, such as sex workers, de facto de-platforming them (Blunt and Wolf, 2020).

Critics of the risk society approach dispute its focus on industrialisation and on modern risks, as well as its downplaying of class and of other risks (Bergkamp, 2017; Curran, 2013). However, considering it has already been used to explain moderation techniques such as shadowbans (Are, 2021a, 2021b, 2021c, 2021d, 2021e), the World Risk Society model can also be applied to automated deletions of profiles posting nudity and sexuality.

A mix of legitimate concerns about harmful behaviours and their deriving moral panics often results in civil society and governments pushing platforms to regulate specific content on social media (Tiidenberg, 2021). To maintain their power, platforms need to be seen to be taking actions against a set of loosely defined and ominous 'online harms', and they often do so by identifying targets, or risks, to contain – risks and targets that, often, have were identified in women's bodies and nudity (Are 2021a, 2021b, 2021c,

2021d, 2021e; Paasonen et al., 2019; Tiidenberg, 2021; Tiidenberg and van der Nagel, 2020).

Such risk-oriented, moral panic-informed governance reflects, for Schoenebeck and Blackwell, a typically Western criminal justice framework, ‘identifying perpetrators of undesirable behavior and administering punitive responses’ (Schoenebeck and Blackwell, 2021: 14). Sanctions such as content or account deletion mirror the problems of retributive models of governance, such as overlooking the needs and interests of targets of harassment while removing offences and offenders from the community without trying to rehabilitate them, and treating all violators equally, so much so that users who unintentionally infringe guidelines are subject to the same sanctions as users who cause harm deliberately (Schoenebeck and Blackwell, 2021). This ostracization of nudity, sex and sex work has been compared to Medieval banishment, an exclusion from public life which is both a consequence of leaving private companies in charge of public space and a reaction to society’s focus on identifying and preventing risks (Are, 2021c; Moeckli, 2016). Furthermore, platforms’ a priori identification of ‘risky’ and ‘problematic’ sexual content often clashes with the legal notion of prior restraint, or a state action prohibiting ‘speech or other forms of expression before they can take place’ (Council of Europe, 2018). Essentially, by often censoring sexual communications before they are even shared by virtue of automatically deleting posts or profiles, platforms’ affordances are causing a chilling effect on sexual communications.

This paper will now showcase instances of ‘automated powerlessness’ faced by accounts posting nudity and sexuality on Instagram and TikTok, accounts who are seen as ‘risky’ and ‘problematic’ just for existing on platforms (Are and Paasonen, 2021).

Methodology

This paper sets out to answer the following research questions:

RQ1: Which in-platform barriers do users posting nudity and sexuality encounter once their profile is deleted from Instagram?

RQ2: Which in-platform barriers do users posting nudity and sexuality encounter once their profile is deleted from TikTok?

Atay (2020) argues that when analysing digital spaces, researchers need fresh new methods. Therefore, I answered my research questions by carrying out a digital autoethnography ‘to describe and systematically analyse personal experience in order to understand cultural experience’ (Ellis et al., 2011: 273). Autoethnography is an ‘interpretation and creation of knowledge rooted in the native context’, using tenets of autobiography and ethnography that, in this case, will be used to describe direct experience of social media moderation’s processes (Mittra, 2010: 15). While this method allows authors to exercise creative licence, Stahlke Wall nonetheless argues that it requires them ‘to be clear about their purpose, provide a level of analysis, and attend to the ethical issues that arise in this form of work’, while also following the pattern of an academic paper, with methodological and findings discussions (Stahlke Wall, 2016: 5). This is how I have

chosen to structure my autoethnography in this paper: as such, the following sections describe my choice of method, its limitations and my positionality, sharing my findings in a narrative format and outlining my purpose and contribution to the field. Digital ethnographies present ‘reflexive, critical, creative, evocative, and poetic first-person narratives’ to examine ‘cultural identity presentations and performances in digital spaces’ (Atay, 2020: 272). Specifically, this paper takes an approach which regards my social media presence itself as a form of digital autoethnography, treating it as a narrative, a form of ‘digital life writing’ which, when threaded together, ‘can tell the theoretically infused stories about the cultures in which we are situated’ (Atay, 2020: 273). I have subconsciously applied this approach as I documented my experiences of social media censorship and content creation, viewing my presence on social media as real-time research in practice since my PhD years. As such, I embraced Ellis & Bochner’s (2006) view of this method as ‘unruly, dangerous, vulnerable, rebellious, and creative’, allowing my experiences of platform governance to take me to places I would have not envisaged – that is being enraged by it (p. 433).

Similarly to autoethnographies outside of the digital realm, I engaged in data collection – in this case, noting down my experiences – *before* finalising my research questions, which were informed by the experiences gathered. Having already conducted and published autoethnographies of my experiences of shadowbanning since 2018 (Are, 2021c), I have continued noting down and writing about the practicalities of censorship arising from running the pole dance and blogging Instagram and TikTok accounts @bloggeronpole (at 23,700 and 345,000 followers respectively as of 26 November 2022). Following my progress from recreational pole dancer to pole dance instructor, my accounts have become well-known in the pole dance industry and have resulted in online and offline job opportunities, such as speaking engagements, workshop teaching, modelling and brand partnerships. Therefore, my Instagram and my TikTok profile have become a source of income as much as a platform to promote my research, my writing, my activism campaigns and my pole dance classes and tutorials, which fully moved online during the UK Covid-19 lockdowns of spring and autumn 2020, and of winter 2021.

Consistently with Banet-Weiser’s (2018) and Duffy and Meisner’s (2022) economy of visibility, which sees social media creators having to constantly produce and publish content to remain relevant and continue working, my pole dancing accounts rely on both digital and physical labour, requiring the effort of creating combinations of pole moves and choreographies, which are then filmed and posted towards the cultivation of platforms that I have been building since 2016. Yet, although social media can be considered an extension of my workplace, it can be argued that, due to their moderation practices, I have found myself either hidden from or outright locked out of my office. This started in 2019, when recreational pole dancers who, like the strippers who popularised our sport (Are, 2021a, 2021b, 2021c, 2021d, 2021e), use dance poles in performances and exercise classes, also became affected by FOSTA/SESTA (Are, 2019b, 2021c). Since pole dance’s aesthetic originates from and mirrors strippers’ one, and given that a form of nudity is necessary for gripping the pole, stripping and pole dancing are often too similar for automated moderation to separate (Are, 2018).

This paper is informed by the experience of running the aforementioned accounts, as well as by questions answered by Instagram press and Facebook Policy teams, and by

TikTok's Creator Support and Creator Communication teams via email. These anonymised responses were published on *bloggeronpole.com*, my fitness, lifestyle and activism blog, launched in December 2017 and currently averaging 15,000 readers per month, with a Domain Authority of 40 (Moz, n.d.; Siteworthtraffic, n.d.). Consistently with my approach to blogging and posting on social media platforms, this paper takes a narrative approach to discuss my experiences of 'automated powerlessness' to answer the research questions.

Since this study mainly features my experiences, it presents a set of limitations. Critics of autoethnography claim examining one's own experience results in researchers being overly immersed in – and not impartial about – their own research (Mitra, 2010). Autoethnography also risks returning subjective findings to the researcher's understanding of a subject, their background and opinion (Mitra, 2010). Thus, it is necessary to acknowledge my positionality: this paper analyses a white, bisexual, cis-gender, thin woman's experiences of social media moderation, meaning that moderation of bodies from a different background may result in different observations. However, the fact that even a white, cis-gender woman pole dancing inside a dance studio can face considerable social media censorship raises important questions on the effects censorship can have on users from less privileged backgrounds, and is therefore worthy of investigation.

I have begun defining my experiences as a 'freak case' in platform governance – an online moderation researcher with unique but replicable experiences of content and account censorship, generating ironic headlines and news stories. Thanks to these experiences, I am in a unique position to discuss falling through the cracks of online moderation. Therefore, precisely because of these experiences and because of the developing status of social media moderation of bodies, autoethnography is an appropriate method to understand how Instagram moderate specific content.

I direct any concerns of bias in the narration shared in this paper to this study's findings, which highlight that platforms are not transparent and do not hold themselves accountable about their processes, whether one interacts with them wearing a user, journalist or a researcher's hat. Therefore, examining this from the perspective of a user with specific experiences becomes valid and applicable to a variety of different user populations.

For all the above reasons, this study makes a step towards understanding Instagram's and TikTok's automated account deletion to highlight the challenges users face when losing their profiles under a system devoid of transparency and accountability.

Implied nudity, inefficient moderation: an 'abusive relationship' with TikTok

My already moderately popular TikTok profile, at 27,000 followers by February 2021, received a boost after one of my videos went viral. A snippet of one of my pole dance choreographies to *WAP*, the hit song by Cardi B and Megan Thee Stallion, the post joked that I was now a pole dancer 'WAP – With A PhD', after my PhD corrections were accepted. I was finally a doctor, and I was keen on sharing it on my social media. In less than a week, the video reached 1 million views and my following went from a manageable 27,000 to an overwhelming 70,000.

However, virality on TikTok has a dark side: as I and a variety of creators told Insider (Perrett, 2021), going viral means appearing on other users' 'For You' page, TikTok's main newsfeed – and many of these users may hate your content. They certainly hated mine, and proceeded to claim I had no place on a platform 'for kids', that my morals were non-existent, that I displayed 'fatherless behaviour' and that I deserved to be raped. While users are not privy to whether their account is being reported, and by how many people if so, it would be fair to say that at least a portion of those displaying such strong reactions to my post may have flagged my account. As a result, after going viral my videos were taken down: I initially lost my posting privileges and was banned from sharing any new content, including a major brand partnership with a popular pole dance brand, for about a week; meanwhile, the abusive comments underneath my videos were not considered in violation of community guidelines.

What followed was a game of whack-a-mole between me and TikTok: as soon as my posting privileges would return and my videos would be restored, something else would be reported and I would lose my posting privileges again. Crucially, content that was flagged and restored would then be taken down again and lost forever – even when moderators previously judged it as not against guidelines. I started stacking up violations, my content started to disappear and, shortly after, my account was deleted.

My frustration, mainly voiced through Instagram and through Twitter, attracted the attention of journalist Chris Stokel-Walker, who interviewed me for an *Input Mag* piece titled: 'TikTok censored a pole-dancing PhD who studies how social media silences women' (Stokel-Walker, 2021). The article resulted in my account and in my posting privileges being restored, as well as in an email from TikTok's UK Communications team apologising for the mistake.

Given the online abuse and censorship I was targeted with, my experience of posting pole dancing content on TikTok was traumatic and disempowering to say the least: not only I could not post and essentially do my job, fulfilling the obligations I had towards the brands I was partnering with and advertising my online dance classes; I was being abused through misogynistic comments and threats throughout, and while these comments were seen as fine by the platform, it was my content (and therefore my earnings) that were being penalised.

The situation worsened in the spring of the same year. Throughout the end of April and in early May, I began experiencing repeated video take-downs and bans from posting, liking or commenting on TikTok. Other users would receive a: 'Please note: This account was reported for multiple community guidelines violation' warning when they followed me or looked at my profile. After my 'violations' started to stack up, I was deleted three times for pole dancing on TikTok in the second week of May alone. Each time, TikTok's Creator Support and their community managers and communication leads would apologise or respond by saying my account did not actually violate their community guidelines. My profile was reinstated. . . only to be deleted shortly after, sometimes within days or hours from the last apology. I appealed each time, but after the third deletion, I received a notification that my account could no longer be recovered.

Having once again contacted both TikTok's communication department and their Creator Support team, I received a partial explanation of the reasons why my account was being constantly deleted: that posting 'implied' or 'partial' nudity would trigger

automated deletions. No further clarification was given, or indeed an explanation of what ‘implied’ nudity means – given that we can imply we are all naked underneath our clothes (Are, 2021e).

I have journalist Chris Stokel-Walker to thank for, once again, helping me recover my account: indeed, I did not recover it because the platform’s infrastructure allowed me to appeal or through speaking by their team, but only because Stokel-Walker took pity on me, mentioned their CEO on Twitter and called their PRs, who called my case ‘mismoderation’ and said they were investigating the issue to stop it happening again.

As a follower told me when I shared my experience, I seem to be in an ‘abusive relationship’ with TikTok. As an abusive relationship survivor, this is scarily accurate: the back-and-forths with the platform, the unclear moderation, the gaslighting originating from this experience are distressing, frustrating and just not an efficient moderation system for either creators or TikTok themselves – particularly when content that is judged to be fine multiple times is deleted after the next viral video.

Without my journalistic contacts and my academic profile, I would have most likely struggled to keep my TikTok profile, and the platform’s lack of communication and transparency may be leaving its many users completely powerless in the face of abuse through commenting *and* through mass reporting.

‘Ass is fine, but we draw the line at grandmas’: losing my main platform after Instagram account deletion

My Instagram account @bloggeronpole and my blog bloggeronpole.com have become a known voice and resource educating about social media censorship and content moderation in my network of pole dancers, sex workers, activists, athletes, artists and educators since 2017. In the space of 2 years, I obtained an apology about shadowbanning from Instagram’s press team in 2019 (Are, 2019a); I was part of the group that created the #EveryBODYVisible protest against Instagram censorship, receiving support by Burlesque superstar Dita Von Teese and acknowledgement from Instagram CEO Adam Mosseri (Are, 2021c); I created a change.org petition against Instagram’s new terms of use, gathering over 122,000 signatures at the time of writing (Are, 2020b) and eventually starting up an ongoing dialogue with Facebook’s policy team (Are, 2021a, 2021b); I have published two academic papers on the Instagram shadowban (Are and Paasonen, 2021; Are, 2021c). In short, I am a known voice in the online moderation activism and research space, a voice Meta are aware of and a voice they communicate with both via email and through videoconferencing.

The fact that my profile was known to Instagram did not spare it from deletion without a warning on 20 July 2021. My last post on my feed had been a picture of me with my 93-year-old grandmother, whom I had met for the first time after being separated from her by lockdowns.

When I logged onto my Instagram profile at 7PM Italian time that day, ready to upload my daily post, I received a notification that my account had been disabled for an unspecified violation of community guidelines. Given that my account has so far been my main tool to express myself, publicise my pole dance classes and tutorials, to raise awareness of pressing issues related to online censorship, to learn from and keep in touch

with my network of pole dancers and activists, I was devastated. I felt lost and completely powerless, conscious that an account I spent almost 10 years building for it to reach nearly 20,000 followers was now gone.

Having received no warning from Instagram – despite its infamous ‘strike’ system warning creators that their account might be deleted due to active community guidelines violations (Kastrenakes, 2019) – I had no way to know what I had done to deserve de-platforming, or if there was anything I could have done to recover my account. All I could do through the appeals system was submit my phone number and email address and wait.

Given my aforementioned ‘freak case’ status I decided to use multiple tactics to attempt to recover my profile. Firstly, I emailed Instagram’s press consultancy in London, even though I knew they would have only replied the following morning as I was emailing after working hours; secondly, I emailed my contacts within Facebook Policy in San Francisco, whom I had been in contact with since December and who had previously helped reactivating accounts from my network that were deleted by Instagram by mistake; thirdly, I started furiously tweeting the impact this deletion was having on my work and on my mental health, since this had been helpful in recovering my TikTok account earlier in 2021. Mainly thanks to my Twitter activity, I attracted the attention of two journalists, from the UK *i Paper* and the *Daily Dot* website, who interviewed me for articles about my experience to be published the next day. Meanwhile, on Instagram, friends and followers directed their audience to my back-up account, a measure many fellow NSFW creators in my network have to take to still be able to communicate when under threat of profile deletion. These users also asked their followers to report my deletion as a mistake under the ‘Something isn’t working’ option in the settings within Instagram’s Help Centre (Are, 2021d).

Only the day after, on July 21, things started moving. I received a response from Instagram’s UK press team, who confirmed they were looking into my deletion. Around 5 PM on that same day, the same person told me that my account would be up and running soon, and that it was deleted in error. Shortly after, the account came back, and Facebook Policy in San Francisco said it was reactivated before their team could check what had happened. However, one of my followers, who worked at Meta and whom I have kept anonymous on the blog, sent me a direct message once I recovered my account saying they themselves had flagged my deletion to their own team, and that while they could not go into too much detail, my account was probably deleted because of a series of ‘false positives’. To this day, I have no real knowledge of what happened to my profile. As I joked on the *TechDirt* podcast, did Instagram’s automated moderation think: ‘Ass is fine, but we draw the line at grandmas?’ (Masnick, 2021).

Either way, what is clear from my experience is that had I not had the platform and connections I had, I would have not been able to recover my account this quickly, or ever, through Instagram’s automated moderation. My experience shows the powerlessness users face when their account is deleted, and the lack of transparency by the platform when it comes to letting creators know what triggered their deletion, and how and if they can get their account back. The complete absence of a ‘deleted creators team’ and the utter lack of communication with the platform makes the deletion of your Instagram profile a crippling, disempowering experience for creators who make their money and

spread their message through the app. This feeling of disempowerment is enhanced by the platform's cryptic responses, and by the feeling that your existence on Instagram can be reversed at the flick of an (automated and devoid of context) switch: you can lose your main source of income in a second, and nobody inside cares about helping you get it back. . . unless, like me, you are a potential threat to their image.

Conclusion: from 'automated powerlessness' towards empowering users

This paper has shared two case studies of automated account deletions on mainstream social platforms TikTok and Instagram, used by creators to both express themselves and conduct their business through advertising their own products or their brand partnerships. Sharing my unique experiences as a 'freak case' – an online moderation researcher with multiple experiences of censorship – I have shown the 'automated powerlessness' triggered by in-platform barriers that users posting nudity and sexuality face once their profiles are deleted by TikTok and Instagram. These barriers, which result in the inability to work and maintain their network in a space that is driven by visibility (Duffy and Meisner, 2022), are created by platforms' own affordances (Graves, 2007), such as:

- deletions triggered by mass flagging by other accounts;
- the inability to appeal when faced by repeated and targeted flagging campaigns and/or mistaken content and account deletions;
- the inability to directly get in touch with platforms about deletions unless the user in question has friends in high places;
- platforms' lack of transparency in informing users about the reasons behind their deletions, preventing them from being able to 'improve' their behaviours on their apps.

'Automated powerlessness' arises when the powers social media afford to a group – for example, platforms themselves, or a coalition of malicious accounts – outweigh the possibilities targeted users have to question or reverse decisions, or to rehabilitate themselves after actions triggered by affordances. In short, an infrastructure and internal processes that allow de-platforming, but not recovery or rehabilitation, are an inadequate form of governance. This is consistent with Schoenebeck and Blackwell's argument that platform governance:

relies on obfuscated processes of content moderation that have little transparency or accountability to all involved parties; content is deleted without leaving any visible trace of its removal; policy violators have little opportunity for recourse and may not even be informed of the specific rule they have broken; reporters receive generalized responses that often don't reference the content in question, if they receive a response at all (Schoenebeck and Blackwell, 2021: 6).

This opacity is particularly evident in cases where malicious reporting might have taken place, since platforms are not compelled to honour the flags they receive, but they

can use them to legitimise the removal of potentially problematic content and to frame the removal as a performance of ‘listening to the community’ (Crawford and Gillespie, 2016).

The problems faced by users posting nudity and sexuality show that platforms like TikTok and Instagram view the presence of users posting nudity and sexuality as inherently problematic and risky, creating power imbalances between those in charge of identifying risks and users not considered worthy of protection and/or existing on their servers (Beck, 2006; Giddens, 1998). As argued by this paper’s adaptation of World Risk Society theory to automated account deletions and confirming Are and Paasonen’s (2021) argument that nudity is, for platforms, problematic as such. And while safety on social media platform is an important issue, it begs the question: safety *for whom?* If we understand safety as protection from harm and the opposite of risk, platforms, their governors and the stakeholders pushing for more content removals too often overlook another key aspect of safety, which is the ability to access and successfully participate in a society which is increasingly becoming digital (Are, 2020a, 2020b, 2020c). Therefore, towards fairer governance, this paper suggests tackling these users’ sense of ‘automated powerlessness’ in the face of online moderation can be mitigated with three updated processes:

- (1) *Improved transparency*: At present, users do not receive direct and specific information about their account’s deletion and/or shadowbanning, meaning they often have to rely on ‘algorithmic folklore’ to explain a phenomenon platforms deny (Cotter, 2021; Savolainen, 2022). By providing a higher, more personalised level of detail about a creator’s account deletion, users can be armed with more information when attempting to appeal platforms’ decision, and can learn to better adhere to platforms’ guidelines;
- (2) *Recognition of malicious flagging as a form of online abuse*: Each time I have discussed my experiences with Instagram and TikTok, I received outright denials that malicious flagging was taking place, when comments underneath my posts proved otherwise. Platforms should recognise creators’ vulnerability to mass flagging and the impact this practice has, particularly on activists and on creators posting nudity and sexuality. This vulnerability should be taken into account when moderating content;
- (3) *Investment in ‘deleted creators’ communications teams*: My inability to reach out to a dedicated Instagram or TikTok moderator to help me recover my account or understand what led to its deletion betrays a lack of care for some of platforms’ main audiences and customers. A team dedicated to supporting creators with more than generalised regurgitations of community guidelines would result in better experiences on platforms and in less instances of account deletions that threaten social networks’ own image.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Carolina Are  <https://orcid.org/0000-0003-1110-3155>

References

- Are C and Paasonen S (2021) Sex in the shadows of celebrity. *Porn Studies* 8: 411–419.
- Arriagada A and Bishop S (2021) Between commerciality and authenticity: The imaginary of social media influencers in the platform economy. *Communication Culture & Critique* 14(4): 568–586.
- Atay A (2020) What is cyber or digital autoethnography? *International Review of Qualitative Research* 13(3): 267–279.
- Are C (2018) Pole dance: Your questions answered. *Blogger On Pole*. Available at: <https://bloggeronpole.com/2018/01/pole-dance-your-questions-answered/>
- Are C (2019a) Instagram denies censorship of pole dancers and sex workers. *Blogger On Pole*. Available at: <https://bloggeronpole.com/2019/07/instagram-denies-censorship-of-pole-dancers-and-sex-workers/> (accessed 5 September 2022).
- Are C (2019b) Instagram apologises to pole dancers about the shadowban. *Blogger On Pole*. Available at: <https://bloggeronpole.com/2019/07/instagram-apologises-to-pole-dancers-about-the-shadowban/> (accessed 5 September 2022).
- Are C (2020a) Instagram quietly admitted algorithm bias . . . but how will they fight it? *Blogger On Pole*. Available at: <https://bloggeronpole.com/2020/06/instagram-quietly-admitted-algorithm-bias-but-how-will-it-fight-it/> (accessed 5 September 2022).
- Are C (2020b) Instagram, stop unfairly censoring nudity and sexuality. *change.org*. Available at: <https://www.change.org/p/instagram-instagram-stop-unfairly-censoring-nudity-and-sexuality> (accessed 5 September 2022).
- Are C (2020c) A corpo-civic space: A notion to address social media's corporate/civic hybridity. *First Monday*. Available at: <https://firstmonday.org/ojs/index.php/fm/article/view/10603/9549> (accessed 5 September 2022).
- Are C (2021a) Petition update: Sharing community experiences of moderation with big tech. *bloggeronpole.com*. Available at: <https://bloggeronpole.com/2021/03/petition-update-moderation/> (accessed 5 September 2022).
- Are C (2021b) Updates on shadowbanning, nudity and ads on Instagram. *bloggeronpole.com*. Available at: <https://bloggeronpole.com/2021/06/updates-on-shadowbanning-nudity-ads-instagram/> (accessed 5 September 2022).
- Are C (2021c) The shadowban cycle: An autoethnography of pole dancing, nudity and censorship on Instagram. *Feminist Media Studies*. Epub ahead of print 19 May 2021. DOI: 10.1080/14680777.2021.1928259.
- Are C (2021d) I was deleted by Instagram. *bloggeronpole.com*. Available at: <https://bloggeronpole.com/2021/07/i-was-deleted-by-instagram/> (accessed 5 September 2022).
- Are C (2021e) Pole dancing on TikTok and the implied nudity ban. *bloggeronpole.com*. Available at: <https://bloggeronpole.com/2021/05/pole-dancing-on-tiktok-and-the-implied-nudity-ban/> (accessed 5 September 2022).
- Banet-Weiser S (2018) *Empowered – Popular Feminism and Popular Misogyny*. Durham: Duke University Press.
- Bartlett J (2018) *The People vs Tech: How the Internet Is Killing Democracy (And How We Save It)*. London: Ebury Digital.
- Beck U (1992) *Risk Society: Towards a New Modernity (Trans. M Ritter)*. London: SAGE.
- Beck U (2006) Living in the World Risk Society. *Economy and Society* 35(3): 329–345.

- Bergkamp L (2017) The concept of risk society as a model for risk regulation – its hidden and not so hidden ambitions, side effects, and risks. *Journal of Risk Research* 20(10): 1275–1291.
- Binns R (2019) Human judgement in algorithmic loops; Individual justice and automated decision-making. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3452030 (accessed 5 September 2022).
- Blunt D and Wolf A (2020) Erased: The impact of FOSTA-SESTA and the removal of Backpage on sex workers. *Anti-Trafficking Review* 14: 117–121.
- Center for Countering Digital Hate (2022) HIDDEN HATE How Instagram fails to act on 9 in 10 reports of misogyny in DMs. Available at: <https://counterhate.com/research/hidden-hate/> (accessed 5 September 2022).
- Clark-Flory T (2019) A troll’s alleged attempt to purge porn performers from Instagram. *Jezebel*. Available at: <https://jezebel.com/a-trolls-alleged-attempt-to-purge-porn-performers-from-1833940198> (accessed 5 September 2022).
- Cotter K (2021) “Shadowbanning is not a thing”: Black box gaslighting and the power to independently know and credibly critique algorithms. *Information Communication & Society*. Epub ahead of print 28 October 2021. DOI: 10.1080/1369118X.2021.1994624.
- Council of Europe (2018) Prior restraints and freedom of expression: the necessity of embedding procedural safeguards in domestic systems. Available at: <https://rm.coe.int/factsheet-prior-restraints-rev25may2018/16808ae88c#:~:text=Prior%20restraints%20are%20State%20actions,before%20they%20can%20take%20place> (accessed 5 September 2022).
- Crawford K and Gillespie T (2016) What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society* 18(3): 410–428.
- Curran D (2013) Risk society and the distribution of bads: Theorizing class in the risk society. *British Journal of Sociology* 64(1): 44–62.
- Diaz A and Hecht-Felella L (2021) *Double Standards in Social Media Content Moderation*. New York: Brennan Centre For Justice.
- Duffy BE and Meisner C (2022) Platform governance at the margins: Social media creators’ experiences with algorithmic (in)visibility. *Media Culture & Society*. Epub ahead of print 23 July 2022. DOI: 10.1177/01634437221111923.
- Ellis C, Adams TE and Bochner AP (2011) Autoethnography: An overview. *Historical Social Research/Historische Sozialforschung* 36: 273–290.
- Ellis CS and Bochner AP (2006) Analyzing analytic autoethnography. *Journal of Contemporary Ethnography* 35: 429–449.
- Fiore-Silfvast B (2012) User-generated warfare: a case of converging wartime information networks and coproductive regulation on YouTube. *International Journal of Communication* 6: 1–24.
- Gerrard Y and Thornham H (2020) Content moderation: Social media’s sexist assemblages. *New Media & Society* 22(7): 1266–1286.
- Giddens A (1998) Risk society: The context of British politics. In: Franklin J (ed.) *The Politics of Risk Society Order*. Cambridge: Polity Press, pp.23–34.
- Gilbert D (2020) Leaked documents show Facebook’s absurd ‘breast squeezing policy’. *Vice*. Available at: <https://www.vice.com/en/article/7k9xnb/leaked-documents-show-facebooks-absurd-breast-squeezing-policy> (accessed 5 September 2022).
- Gillespie T (2010) The politics of platforms. *New Media & Society* 12(3): 347–364.
- Gillespie T (2018) *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University.
- Graves L (2007) The affordances of blogging: A case study in culture and technological effects. *Journal of Communication Inquiry* 31(4): 331–346.

- Haimson O, Delmonaco D, Nie P, et al. (2021) Disproportionate removals and differing content moderation experiences for conservative, transgender, and black social media users: marginalization and moderation gray areas. *Proceedings of the ACM on Human-Computer Interaction*, 5: 1–35.
- Hudson B (2003) *Justice in the Risk Society: Challenging and Re-affirming Justice in Late Modernity*. London: SAGE.
- Instagram (n.d.a) @bloggeronpole. Available at: [instagram.com/bloggeronpole](https://www.instagram.com/bloggeronpole) (accessed 5 September 2022).
- Instagram (n.d.b) Help center. Available at: <https://help.instagram.com/477434105621119> (accessed 5 September 2022).
- Jane EA (2014) ‘Back to the kitchen, cunt’: Speaking the unspeakable about online misogyny. *Continuum* 28(4): 558–570.
- Kastrenakes J (2019) Instagram will now warn users close to having their account banned. *The Verge*. Available at: <https://www.theverge.com/2019/7/18/20699393/instagram-account-ban-warning-message-moderation-update> (accessed 5 September 2022).
- Kaye D (2019) *Speech Police – The Global Struggle to Govern the Internet*. New York, NY: Columbia Global Reports.
- Masnick M (2021) The impact of shadowbanning. *TechDirt*. Available at: <https://soundcloud.com/techdirt/the-impact-of-shadowbanning> (accessed 5 September 2022).
- Mitra R (2010) Doing ethnography, being an ethnographer: The autoethnographic research process and I. *Journal of Research Practice* 6(1): 1–21.
- Moeckli D (2016) *Exclusion From Public Space: A Comparative Constitutional Analysis*. Cambridge, MA: Cambridge University Press.
- Moz (n.d.) Available at: <https://moz.com/link-explorer> (accessed 5 September 2022).
- Norman DA (1988) *The Psychology of Everyday Things*. New York, NY: Basic Books.
- Paasonen S, Jarrett K and Light B (2019) *#NSFW: Sex, Humor, and Risk in Social Media*. Cambridge, MA: MIT Press.
- Peake E (2022) Instagram, nudes and Tommy Lee: Why the feminist shadow-banning fight is far from over. *iNews*. Available at: <https://inews.co.uk/inews-lifestyle/instagram-nudes-tommy-lee-feminist-shadow-banning-women-censorship-facebook-meta-1808086> (accessed 5 September 2022)
- Perrett C (2021) Transgender TikTok creators say the app’s mysterious “For You” page is a breeding ground for transphobia and targeted harassment. *Business Insider*. Available at: <https://www.businessinsider.com/tiktok-transphobia-problem-creators-report-harassment-threats-2021-2?r=US&IR=T> (accessed 5 September 2022).
- Peterson C (2013) *User-Generated Censorship: Manipulating the Maps Of Social Media*. MIT Graduate Program in Comparative Media Studies. Available at: <https://cms.mit.edu/user-generated-censorship/>
- Sablosky J (2021) Dangerous organizations: Facebook’s content moderation decisions and ethnic visibility in Myanmar. *Media Culture & Society* 43(6): 1017–1042.
- Salty (2019) Exclusive: Victoria’s Secret influence on INSTAGRAM’S censorship policies. *Saltyworld.net*. Available at: <https://saltyworld.net/exclusive-victorias-secret-influence-on-instagram-censorship-policies/> (accessed 5 September 2022).
- Savolainen L (2022) The shadow banning controversy: Perceived governance and algorithmic folklore. *Media Culture & Society* 44: 1091–1109.
- Schoenebeck S and Blackwell L (2021) Reimagining social media governance: Harm, accountability, and repair. *Yale Journal of Law and Technology* 23: 113–152.
- Siteworthtraffic (n.d.) bloggeronpole.com. Available at: <https://www.siteworthtraffic.com/> (accessed 5 September 2022).

- Sloan L and Quan-Haase A (2017) *The SAGE Handbook of Social Media Research Methods*. London: SAGE Publications.
- Stahlke Wall S (2016) Toward a moderate autoethnography. *International Journal of Qualitative Methods* 15: 1–9.
- Stefanello V (2021) Instagram bans popular pole-dancing blogger and activist, who says now “no one is safe” with its sex policies. *The Daily Dot*. Available at: <https://www.dailydot.com/debug/instagram-ban-pole-dancer-bloggeronpole/> (accessed 5 September 2022).
- Stokel-Walker C (2021) TikTok censored a pole-dancing PhD who studies how social media silences women. *Input Mag*. Available at: [https://www.inputmag.com/culture/tiktok-censored-banned-pole-dancer-phd-\[carolina-are\]](https://www.inputmag.com/culture/tiktok-censored-banned-pole-dancer-phd-[carolina-are]) (accessed 5 September 2022).
- Suri S and Gray ML (2019) *Ghost Work: How to Stop Silicon Valley From Building a New Global Underclass*. Boston, MA: Houghton Mifflin Harcourt.
- Tiidenberg K (2021) Sex, power and platform governance. *Porn Studies* 8(4): 381–393.
- Tiidenberg K and van der Nagel E (2020) *Sex and Social Media*. Melbourne, VIC: Emerald Publishing.
- TikTok (n.d.) Adult nudity and sexual activities. *Community Guidelines*. Available at: <https://www.tiktok.com/community-guidelines?lang=en#30> (accessed 5 September 2022).
- Vivienne S (2016) *Digital Identity and Everyday Activism – Sharing Private Stories With Networked Publics*. Basingstoke: Palgrave MacMillan.
- Zuckerberg M (2018) A Blueprint for content governance and enforcement. *Facebook*. Available at: <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/> (accessed 5 September 2022).