Introducing Vigilant Audiences

This ground-breaking collection of essays examines the scope and consequences of digital vigilance—a phenomenon emerging on a global scale, which sees digital audiences using social platforms to shape social and political life. Longstanding forms of moral scrutiny and justice seeking are disseminated through our contemporary media landscape, and researchers are increasingly recognising the significance of societal impacts effected by digital media.

The authors engage with a range of cross-disciplinary perspectives in order to explore the actions of a vigilant digital audience—denunciation, shaming, doxing—and to consider the role of the press and other public figures in supporting or contesting these activities. In turn, the volume illuminates several tensions underlying these justice-seeking actions—from their capacity to reproduce categorical forms of discrimination, to the diverse motivations of the wider audiences who participate in vigilant denunciations.

This timely volume presents thoughtful case studies drawn both from high-profile Anglo-American contexts, and from developments in regions that have received less coverage in English-language scholarship. It is distinctive in its focus on the contested boundary between policing and entertainment, and on the various contexts in which the desire to seek retribution converges with the desire to consume entertainment.

As with all Open Book publications, this entire book is available to read for free on the publisher’s website. Printed and digital editions, together with supplementary digital material, can also be found at www.openbookpublishers.com.
Doxing, the public release of personally identifiable information, is both a means of harassment and intimidation, and a tool for activism. Releasing personal information about individuals whose actions or stated beliefs harm others, or undermine social cohesion, removes the anonymity or obscurity that may foster these forms of antisocial behaviour. However, doxing may also be used to target those whose actions or beliefs pose no risk to others, or who merely do not follow social norms. In societies that value individualism (such as Western liberal democracies), anonymity and obscurity are important means of protecting individuality from the pressures of conformity and popular opinion, and allow us to develop ourselves and experiment with ways of living. Depriving someone of the protections anonymity and obscurity grant them should not be taken lightly. For doxing to effectively support activist goals, it must expose wrongdoing that the broader population (the activist’s audience) will accept as a legitimate concern. Otherwise, activists risk creating sympathy for their target and ostracising themselves from the audience they wish to persuade.

Other chapters in this volume by Mojca M. Plesničar and Pika Šarf, and Tara Milbrandt analyse specific instances in which doxing was used by activists to shame those who use hate speech on social media, or who were present at a white supremacist rally, respectively. This chapter supplements these discussions by presenting a moral justification for doxing as audience vigilantism in response to hate speech. It draws on Daniel Trottier’s (2017) account of digital vigilantism and Emma

© David M. Douglas, CC BY 4.0 https://doi.org/10.11647/OBP.0200.10
Jane’s (2017) analysis of feminist “digilantism” in response to Internet harassment and sexist hate speech. I argue that doxing that deanonymises a proponent of hate speech is an appropriate means of combating hate speech if it is intended to begin a process of deradicalisation.

Identifying sources and promoters of hate speech via doxing serves several purposes. Firstly, it increases the risk of engaging in hate speech by removing the speaker’s anonymity and increasing their vulnerability to legal sanctions and social ostracism. Secondly, it draws attention to threatening behaviour that might otherwise be ignored. Finally, it may serve as a response by the vilified group, or their supporters, that they do not trust the authorities to adequately respond to acts of hate speech against them. Doxing may also be employed to delegitimise public officials who are either themselves anonymous sources of hate speech, or who support it, since promoting hate speech is incompatible with the duty of service to all citizens that public service requires.

This defence of doxing has a number of limitations. As it is a response to the tension between tolerance and freedom of speech, it is intended to apply to liberal democracies that value both. It also does not justify releasing any sort of personally identifiable information (that is, information about someone that can uniquely distinguish them from other people) about practitioners of hate speech. It uses an earlier analysis of doxing (Douglas, 2016) that distinguishes between doxing that is intended to remove someone from obscurity or anonymity (deanonymising doxing), doxing that make it easier to physically locate someone (targeting doxing) and doxing that reveals personal information that undermines the identified person’s credibility (delegitimising doxing). Targeted doxing should also be avoided as the response risks going beyond what Kelly D. Hine calls “socially tolerable bounds” (1998, p. 1253) in her defence of vigilantism. In this case, targeted doxing creates the opportunity for vigilantes to physically intimidate or harm the identified individual. Outside of the potential harm caused to the targeted individual, this may undermine broader support for acts against hate speech.

I also acknowledge that unlike the cases of sexist hate speech Jane describes (in which the audience responds to specific messages by named individuals who do not disguise their identity), there is a significant risk of false positives, in which innocent individuals are erroneously identified
as sources or promoters of hate speech. The negative consequences of mistaken identification mean that those who have erroneously identified an innocent person must take responsibility for correcting this error if this form of deanonymisation is to be permissible.

I will support this argument with short descriptions and discussions of two examples where doxing has been used to either identify proponents of hate speech, or where it is used by hate speech proponents to intimidate activists working against them. I also use the examples of sharing racist speech found on Facebook on the Zlovenija Tumblr page and identifying the participants of the “Unite the Right” rally that are described in other chapters by Plesničar and Šarf, and Milbrandt, respectively. Given the important of hate speech to my defence of this form of audience vigilantism, that is where I will begin my discussion.

**Hate Speech**

Raphael Cohen-Almagor (2015, p. 148) defines hate speech as “bias-motivated, hostile, malicious speech aimed at a person or a group of people because of some of their actual or perceived innate characteristics”. It may also be described as group libel or group defamation (Waldron, 2012, p. 39). More generally, Jeremy Waldron (ibid., p. 4) describes hate speech as something undermining the public good that individuals in a society, from diverse backgrounds and belonging to different groups, will be able to live their lives without facing “hostility, violence, discrimination, or exclusion by others”. It undermines the dignity of those singled out by hate speech by attacking their status as equal members of the community (ibid., p. 5). On a more visceral level, it provokes memories of historical oppression that continue to resonate with those whose identities have been marked as inferior in the past, and who still face discrimination today. In Charles R. Lawrence III’s powerful words:

> There is a great difference between the offensiveness of words that you would rather not hear — because they are labeled dirty, impolite, or personally demeaning — and the injury inflicted by words that remind the world that you are fair game for physical attack, evoke in you all of the millions of cultural lessons regarding your inferiority that you have so
Introducing Vigilant Audiences

painstakingly repressed, and imprint upon you a badge of servitude and subservience for all the world to see. (1990, p. 461, emphasis in original)¹

The continued existence of implicit and explicit biases against minorities (including hate speech itself) suggests that even well-intentioned states fail to practically reflect their stated commitments to political and social equality. Waldron (2012, p. 31) is right in saying that “the position of minority groups as equal members of a multi-racial, multi-ethnic, or religiously pluralistic society is not something that anyone can take for granted”. The targets of hate speech may also believe that the state (or those implementing and enforcing the state’s laws) will not take their concerns seriously (Schwencke, 2017).

The legal tolerance for hate speech differs between societies, even among those recognising the importance of freedom of expression (Sumner, 2003, pp. 144–45). The United States places a strong emphasis on the importance of freedom of expression, and so permits hate speech that would be illegal in other liberal democracies (Cohen-Almagor, 2015, p. 205; Mill, 2018). Even in liberal democracies that criminalise hate speech (such as Australia and the UK), the extent of the limitations they impose on freedom of expression is contested (Mill, 2018).

The targets of hate speech and those who regard it as a serious threat to social cohesion often resort to private measures (including vigilantism) to counter the expressions of hate they face. Such actions are self-defence by the targeted group, and serve two purposes: silencing the hate speech itself (the retributive purpose), and signalling to the broader community that such speech is unacceptable (the expressive purpose). Popular support for the vigilante’s actions depends on the vigilante’s justifications for why suppressing this instance of hate speech is more important than tolerating the speaker’s right to express it.

Digital Vigilantism (DV) and ‘Digilantism’

Trottier describes digital vigilantism (DV) as a response to a transgression that “seek[s] to render a targeted individual (or category of individual) visible through information sharing practices such as assembling and

¹ I originally came across this quote in Hate Crimes in Cyberspace by Danielle Keats Citron (2014, p. 17).
Doxing as Audience Vigilantism against Hate Speech

publishing their personal details” (2017, p. 57). It is made possible by computer-mediated communication systems that facilitate informal groups in coordinating actions in response to transgressions, and dispersing once satisfied that their goal is achieved (ibid.).

Jane’s concept of “digilantism” overlaps with Trottier’s account of DV, and further shows how these activities differ from traditional vigilantism. Jane defines digilantism as “politically motivated (or punitively politically motivated) practices outside of the state that are designed to punish or bring others to account, in response to a perceived or actual dearth of institutional remedies” (2017, p. 3). This description goes beyond the traditional account of vigilantism described by Les Johnston (1996) by incorporating responses that might otherwise be classified as “activism” (Jane, 2017, pp. 3–4). Violence or threatened physical violence usually appear to be a necessary part of vigilantism (Dumsday, 2009; Hine, 1998, pp. 1248–9). Without physical violence (or the threat of it), vigilantism may appear to be just particularly robust activism. Nonetheless, the connection between vigilantism and violence may be maintained by broadening the concept of violence as “avoidable insults to basic human needs” (Gatlung 1990, p. 292). While Johan Gatlung uses this definition to underpin his conceptions of structural and cultural violence (which describe the processes of exploitation and the symbols and language that justify both direct physical violence and exploitation, respectively), this definition may be used to describe vigilantism as violence (either direct, physical violence, or the structural or cultural violence traditionally thought of as “non-violent”) in response to direct, structural or cultural violence by the targets of such violence.

Hine (1998, p. 1222) mentions citizens confronting and harassing drug dealers to force them to leave their community, and residents threatening to launch civil law suits against property owners to compel them to evict criminals, as examples of non-violent vigilantism. Such “non-violent” vigilantism should not be criminalised, Hine (ibid., p. 1252) argues, as the tendency to over-punish transgressors using physical violence makes vigilantism costly to society, rather than vigilantism itself. This allows Hine’s concept of vigilantism without physical violence to fall within the scope of DV and digilantism.

A major form of DV is “weaponising” visibility through “naming and shaming” transgressive individuals or groups (Trottier, 2017, p.
Those targeted by DV are removed from obscurity by the public exposure of a transgression (or alleged transgression). The power of such shaming comes from both the audience witnessing it, and the identified individual’s knowledge that there is such an audience, who may themselves participate in further shaming them. In DV and vigilantism, these are mediated audiences: they witness transgressions through computer-mediated communication systems, such as social media platforms. Furthermore, the audience can increase the visibility of the transgression by sharing and commenting on it on social media. As such, the vigilante audience need not be limited to the same geographical area or country as the targeted transgressor. Transgressions visible and provocative enough to provoke widespread disgust may face a global vigilante audience. Controversial, poor-taste or merely misunderstood social media posts are examples of such transgressions.

**Doxing**

Doxing is the deliberate release of personally identifiable information in a form that is easily accessible to others, usually with the aim of intimidating the identified person (Hawley, 2019, p. 201; Douglas, 2016, p. 199). Joan Donovan (2017) rightly notes that doxing offers “a powerful leveller for those who seek social justice when they know criminal justice is far out of reach”. Using doxing to draw attention to wrongdoers is a form of private justice that comes under the umbrella of DV (Trottier, 2017, p. 56).

The information released by doxing are forms of identity knowledge about specific individuals. Elsewhere I distinguish between three types of doxing: deanonymising, targeting and delegitimising (Douglas, 2016, pp. 203–6). These types differ in what information is revealed about someone and the motive for doing so. Deanonymisation reveals information that connects an individual’s anonymous or pseudonymous identity to their regular identity. Targeting reveals information about an individual that allows them to be physically located, such as their home address or workplace. Delegitimising reveals potentially embarrassing or

---

2 Daniel Solove vividly calls this the “digital scarlet letter” (2007, p. 76).
humiliating information about an individual. Each kind of doxing takes something from the identified person: anonymity (deanonymisation), obscurity (targeting) or credibility (delegitimisation) (ibid., p. 204). Recognising the different types of doxing is important to determine whether there is a plausible moral argument for releasing personal information.

Seeking anonymity is not necessarily a sign of wrongdoing. Gary T. Marx (1999) lists fifteen common rationales for anonymity (or pseudonymity) that are socially desirable. The purpose of seeking anonymity must be considered if deanonymisation is to be justifiable. There are clear cases for justifying deanonymisation when anonymity is adopted to avoid accountability for wrongdoing, and when a pseudonym is used to mislead others (for example, by using an alias to associate themselves with a particular group for some benefit). In these cases, doxing is a form of whistleblowing.

Deanonymisation only requires revealing information that establishes a connection between a pseudonym or actions performed anonymously, and some form of identity knowledge of an individual. One form of identity knowledge is an individual’s locatability, which includes both their physical location and the means through which they can be contacted (ibid., p. 100). Revealing locatability information about an individual increases their vulnerability to harassment (if contact information is disclosed) and physical harm (if location information is disclosed).

If deanonymising and targeting doxing answer the ‘who’ and ‘where’ questions about an identity (ibid., p. 101), delegitimising doxing reveals ‘why’ this particular individual is of interest. It purports to reveal a transgression by the identified individual for which they deserve (in the view of those doxing them) to be publicly shamed. These transgressions may be breaches of social norms, immoral activities, deceptive or harmful behaviour, or anything else that may undermine the individual’s reputation or publicly humiliate them. The transgression is from the perspective of those performing the doxing: the identified individual or the wider community may not recognise it as particularly troubling, threatening, controversial or transgressive at all.
Doxing in Response to Hate Speech

I consider four examples where doxing is used to either identify members of groups that promote hate speech, or members of groups that protest hate speech. These examples are: Anonymous’ #OpKKK campaign to reveal the identities of Ku Klux Klan members; the display of racist speech found on Facebook, posted with the author’s identity on the Zlovenija Tumblr page; the identification of individuals pictured taking part in the 2017 “Unite the Right” rally at Charlottesville, Virginia in the US; and attempts by right-wing groups to identify participants in Antifa protests. In each case, individuals or groups attempt to uncover and disclose identity knowledge (such as the common or legal names) of individuals. In the case of #OpKKK, the hacktivist collective Anonymous sought to reveal the membership of the Ku Klux Klan so that they could not conceal their racist beliefs and activities. In the Zlovenija example, the anonymous activists were naming individuals who expressed hate speech on social media. In the case of identifying participants in the “Unite the Right” rally, individuals sought to remove the obscurity of those who promoted hate speech in a public space. The final case, attempts by right-wing groups to identify members of Antifa groups, differs from the others in that doxing is being used against those who protest against hate speech. It serves as a contrast with the other cases in order to establish that the context within which doxing is used is significant for whether a group may use it to defend itself.

Anonymous and #OpKKK

The Ku Klux Klan (KKK) are an American white supremacist group that has existed in various forms since around 1866 (Law, 2009, p. 128). While its activity and membership has waxed and waned since the group’s emergence, its long history of violence and the distinctive costumes that conceal its members’ identities have made the KKK particularly notorious in the United States. It currently exists as a collection of localised groups that share the name, rituals and the white supremacist ideology of the original group (Southern Poverty Law Center, n.d.).

The impetus for #OpKKK was the distribution of leaflets by a KKK chapter threatening to use “deadly force” against those protesting the
police shooting of Mike Brown in Ferguson, Missouri, on August 9, 2014 (Gilbert, 2014). Activists identifying themselves with Anonymous hacked two Twitter accounts connected to the KKK, and claimed to have gathered information about their members through their access to these accounts (Woolf & Stafford, 2015). Around this time, news reports from left-wing web sites stated that several US senators and mayors were on Anonymous’ list of KKK members (Anonymous OpKKK, 2015). A Twitter account associated with Anonymous distanced itself from these reports (ibid.). A list of around 350 names was eventually released by those involved in #OpKKK (Anonymous, 2015). Some of those listed had already publicly stated their membership of the KKK (Woolf, 2015). The document listing the names of alleged KKK members contains a preface describing the activists’ methods of data collection and a thoughtful account of their reasons for releasing this data (Anonymous, 2015).

Zlovenija

As Plesničar and Šarf describe in their chapter, Zlovenija was a Tumblr page where the anonymous author(s) posted screen captures of racist speech that related to migration to Slovenia. Examples of hate speech were removed if the speaker apologised. The anonymous authors of Zlovenija removed the examples of racist speech after posters displaying the posts featured on the site began to appear on the streets in Ljubljana and due to concerns about organised attempts to deanonymise the author(s).

Identifying “Unite the Right” Rally Participants

The August 2017 “Unite the Right” white-nationalist rally in Charlottesville, Virginia and the digital vigilantism that occurred in response are described in detail by Tara Milbrandt in another chapter of this book. For the purpose of this chapter, I will focus on two aspects of the doxing that happened to rally participants: whether the response was motivated by an acceptable goal, and the instances of mistaken identification that occurred.

This example differs from #OpKKK as the targets are limited to a particular group: those present at the Charlottesville rally in support
of white nationalism. The identified individuals also made no attempt to conceal their identities at the rally. The @YesYoureRacist Twitter account also only listed the names of the identified individuals, making it deanonymising doxing (as they were removed from obscurity). It was also delegitimising doxing as it portrayed the identified individuals as white nationalists. It was not targeting doxing, as no locatability knowledge was revealed by the account.

Identifying Antifa Protesters
Anti-fascists (or Antifa) seek to disrupt the ability of fascist and far-right groups to organise and communicate their views to the public (Bray, 2017, pp. xiv–xv). Doxing perpetrators of hate speech (those creating or spreading hostile, biased and malicious works targeted at specific groups of people) is a controversial method of combating the support and propagation of such speech. For example, doxing is one tactic adopted by Antifa groups to silence members of fascist and far-right groups (ibid., pp. 86–87), although not all Antifa groups accept it as a legitimate tactic (Bartlett, 2015, p. 67). Similarly, fascist and far-right groups themselves use doxing to intimidate critics (such as journalists) (Wilson, 2018) and to expose Antifa members (Bartlett, 2015, p. 68).

Unlike the previous examples, this case covers instances where doxing is used against those who vigorously oppose hate speech and the right-wing groups who are at least sympathetic to it. This example is important for establishing that the activist’s cause in doxing others is important for determining whether it can be defended as socially useful vigilantism. In other words, whether doxing opponents is self-defence for the activists themselves, or self-defence for both the activists themselves and the broader community.

Doxing as Audience Vigilantism
Doxing as a tool for audience vigilantism has three stages: uncovering personally identifiable information, releasing and announcing that information, and the audience acting on that information. For simplicity, I will call these the discovery, release and response stages of doxing as audience vigilantism. The discovery stage may be performed in secret
by activists, while the release and response stages require audiences who serve as the potential vigilantes who will act against the identified individual. How the information is disseminated and presented in the release stage affects the potential audience, the likelihood that the audience will respond and how that audience will harass or shame the identified individual.

The discovery stage is performed by the activist, who has identified a transgressor (or potential transgressor). The activist uses whatever information is available to them to search for further identity knowledge about the transgressor. The activist may seek assistance in identifying an individual by releasing whatever information they have and asking others for further information. In this case, the discovery stage merges with the release and response stages.

The release stage has two components: the publication of identity knowledge about the targeted individual(s), and publicising the release of this information. Gaining publicity for the information release is necessary to create an audience who may act on it. Social media can serve a role, as it does for the discovery stage: in addition to providing evidence of the target’s transgressions, it also offers an avenue for publicising the release of identity knowledge. In the #OpKKK example, the release stage was Anonymous’ posting of information onto Pastebin (Anonymous, 2012) and announcing this on Twitter. In the Zlovenija example, the release stage was the posting of the racist speech and the identity of the speaker on the Zlovenija page. In the “Unite the Right” example, this stage was the posting of pictures of rally participants on the @YesYoureRacist Twitter account, which served as both publishing and publicising the release of this identity knowledge.

The response stage is the actual vigilante action the audience performs. It is the activist’s goal in discovering and releasing information about the target. The audience may respond by further publicising the transgression and the transgressor’s identity, or by using the identity knowledge to contact the transgressor. The first response seeks to shame the identified individual, while the second harasses or threatens them (or worse). The second type of response is traditional vigilantism, as it includes the threat of violence (Johnston, 1996). DV and digilantism only necessarily involves increasing the target’s visibility, and so may include acts where only the first response occurs, or where both the first and
Introducing Vigilant Audiences

the second occur. Ethical evaluations of vigilantism (such as Dumsday (2009)) may therefore be applied to cases that involve both the first and second response. As the literature on traditional vigilantism may be directly applied to the second type of response, my discussion will focus on the first type, where shame is used to punish the transgressor.

In both types of response, the audience’s active involvement takes the response beyond the activist’s control. The appearance of posters around Ljubljana containing posts from the Zlovenija page is an example of how the audience’s response can differ from what the activist anticipates. As Jane (2017, pp. 5–6) rightly notes, diligentism (and social media activity generally) are characterised by “[c]haos, speed, unpredictability, strong affect, and spur-of-the-moment decisions”. This unpredictability and lack of control makes audience vigilantism risky to initiate, and serves as a prima facie moral objection against it as the response is likely to be disproportionate to the transgression.

The #OpKKK example demonstrates the care necessary in performing the discovery and releases stages of doxing as DV, and the external factors that can compromise its effectiveness. The legitimate #OpKKK list described in broad terms how they collected the listed data. The decentralised and anarchic nature of Anonymous creates difficulties for the audience in clearly distinguishing between the ‘legitimate’ Anonymous #OpKKK list and the independent list. The apparent source of the first list did not claim to belong to Anonymous or to be connected with #OpKKK (Woolf & Stafford, 2015). This nuance and the confusion caused by the false initial reports of politicians appearing on the list reduced the impact that the release might otherwise have had. While Anonymous’ unique character made it particularly difficult for the audience to identify ‘official’ Anonymous information releases from others, this problem exists for any activist group. An effective and responsible DV campaign therefore needs to present a clear identity to avoid confusion within the audience it seeks to inform and motivate.

Jane (2017) notes that Hine’s account offers two criteria for vigilantism that should not carry criminal liability. These criteria also suggest how we might begin to morally evaluate DV. These criteria are that the vigilante’s actions are based “on an accurate perception of social need”, and that the act is within “socially tolerable bounds” (Hine, 1998, p. 1952). These criteria suggest some constraints on the possible motives
and actions that might be defended as morally permissible DV. The first, an accurately perceived social need, limits the permissible causes that may motivate DV. These causes must be reasonable in the sense that they can be justified to those with other perspectives within their community (Stanley, 2015, p. 108). This rules out conspiracy theories and extremist political, racial or religious causes, as they cannot be described and defended using premises and evidence that are in principle justifiable to all. Following Hine, I will call this the legitimate social need criterion. This restriction on extremes also carries over to the second criterion, which is that the digital vigilante acts must not cause physical harm to persons, and the disruption caused must be temporary and limited to those responsible for the concern identified in the first criterion. Permissible DV cannot be indiscriminate, and must respond to injustices that are, in principle, recognisable to those unaffected by it. Again following Hine, I will call this the socially tolerable bounds criterion for DV.

The emphasis these criteria place on DV to be understandable and, in principle, acceptable to the broader community highlights how DV (and doxing as DV in particular) is audience vigilantism. DV’s effectiveness in addressing the social concerns that motivate it depends on how well it presents its cause to the audience. Without audience support, this form of vigilantism cannot achieve its intended purpose. Extreme methods and political causes risk alienating audiences and creating sympathy and community support for the targets of DV. Activists who utilise audience vigilantism therefore have an interest in keeping their actions within limits to encourage and maintain popular support for their goals.

Shaming as Audience Vigilantism

Audience vigilantism replaces the physical violence of traditional vigilantism with the shaming of individual(s) who have transgressed against the community. Jacob Rowbottom (2013, p. 1) lists three goals for “naming and shaming” individuals: informal punishment, informing the public about their conduct and expressing disapproval for that conduct.

The use of shame as a legal punishment for wrongdoing offers a useful starting point for considering it as a vigilante punishment.
Shame punishments express that the punished is a certain sort of person (Nussbaum, 2004, p. 230). Shame punishments for hate speech, for example, express that the punished individual is a bigot with a hatred of a certain group of people. The person is made visible as someone of poor character, and so shame punishments risk permanently staining the identity of those they punish (ibid., p. 231). In contrast, guilt punishments express only that the punished has performed a wrongful act (ibid., p. 230). Nussbaum (ibid., p. 207) describes guilt as “a type of self-punishing anger, reacting to the perception that one has done a wrong or a harm”. Guilt punishments condemn the act, while shame punishments condemn the person.\(^3\)

This connection between shame and guilt suggests how shame punishments might be employed without necessarily alienating the transgressor from respectable society. Brooks (2008, p. 330) argues that shame punishments can be justified if used to inspire guilt within the wrongdoer. This moves the purpose of shaming from humiliation to reintegration: the individual’s stigmatisation is temporary until they themselves recognise the wrongness of their actions. Reintegrative shaming follows the transgressor’s shaming with efforts to return them to respectable society through limiting the time they are shamed, offering forgiveness for their actions and recognising that the transgression should not define the person’s identity (Braithwaite, 1989, pp. 100–1). The purpose is to move from condemning transgressors themselves (they are being shamed by others for wrongdoing) to condemning their actions (they themselves feel guilt for wrongdoing, and this guilt can motivate them to change their behaviour). It should encourage them to change their behaviour and attitudes, and encourage their community to forgive their transgression, rather than permanently alienate them from it.

While these discussions of shame punishments within a legal system suggest that shaming may be justified as a punishment, its use by vigilantes introduces problems that make it more difficult to justify. Solove (2007, pp. 94–98) lists several concerns about Internet shaming: its permanence, its disproportionality, the lack of due process and its

---

\(^3\) Zlovenija is an example of this distinction. As the authors removed the post if the person who made the racist comment apologised, they inspired guilt in the speaker without permanently shaming them.
abuse for bullying and personal disputes. The permanence of Internet shaming (due to the accessibility of the shaming material via search engines, and the possibility that others have copied and redistributed it even if the original poster removes it) makes returning to obscurity (effectively reintegrating into society) difficult after being shamed. This is one of the motivations of the “right to be forgotten” (Jones, 2016, pp. 9–15).

The widespread accessibility of the humiliating material also makes Internet shaming particularly harmful. It lacks the definite end of reintegrative shaming, and the transgression may continue to define that person in their interactions with others. The widespread integration of the Internet into daily life makes it difficult to distance oneself from it to avoid encountering harassment (Franks, 2012, p. 682).

The risks associated with shame punishments imposed by vigilantes make them difficult to defend without considering their context. For those unable to seek legal action against legitimate threats to themselves and their communities, and who lack other effective alternatives to defend themselves against aggression, shaming punishments may be justified provided that they are used carefully. The shaming of hate-speech proponents should be directed towards encouraging them to reconsider their views and reintegrate into the broader community. Excessive shaming and public humiliation risk being counter-productive, as social support (and the loss of it) is an important factor in holding and rejecting radical beliefs (Koehler, 2017, p. 17). The concept of reintegrative shaming also suggests how shaming by a vigilante audience might be kept within “socially tolerable bounds”. Defending themselves against hate speech also serves as an ‘accurately perceived social need’ for minority groups.

#OpKKK certainly fulfils the legitimate social need criterion for DV: the KKK exists, it has a history of violence and the fact that some members are comfortable enough to publicly claim to be members suggests that legal responses to the hate speech it promotes are insufficient. Publicising the membership of the KKK draws attention to the threat the organisation poses to others within US society. What of the socially tolerable bounds criterion? Ideally, those identified as KKK members would go through a process of reintegrative shaming, which may involve publicly renouncing their membership and assisting law
Introducing Vigilant Audiences

enforcement agencies with information about criminal activity performed by the organisation. This, of course, carries personal risk for such individuals, but it serves as a clear public signal of guilt about their former harmful transgressions. The community would also be obliged to protect those who have renounced their transgressions from reprisals. The former transgressor and the community are then reconciled through their newfound mutual assistance: the community through reforming someone whose views and actions were harmful to others and obtaining useful information to prevent further hate crimes and hate speech, and the transgressor through protection against reprisals and gaining a more balanced and nuanced perspective on the society in which they live.

The Zlovenija page was motivated by a concern about the racism within public discussions of migrants arriving in Slovenia. It had a legitimate social need as a formal attempt to combat the hate speech found on social media. It also arguably meets the tolerable social bounds criterion. While an argument can be made that including a direct link to the speaker’s Facebook post increased the risk of targeting doxing occurring (since it increased the ease with which location information about the individual could be found if it is included in the person’s profile), the Zlovenija authors acted responsibly by removing the identity information they posted once they became concerned that they were losing control over how this information was being used. Their removal of posts after the speaker apologised also demonstrates the authors’ responsibility in seeking to limit the shaming of those they identified.

The violence and hate speech at the “Unite the Right” rally provide clear evidence that there is a legitimate social need to defend the community against white nationalism. The apparent reluctance of law enforcement to intervene and keep protesters and counter-protesters apart (Thompson, 2017), and the presence of citizen militias whose sympathies were unclear (Gunter & Hughes, 2018) might also be used to argue that the legal means of protecting the community are insufficient and that audience vigilantism is necessary.

These limitations assist in meeting the tolerable social bounds criterion. However, revealing any form of identity knowledge makes it easier to find locatability information about an individual, so the fact
that @YesYoureRacist did not itself reveal locatability information does not mean that those identified could not be targeted (Ellis, 2017). The relative anonymity of participating in a large public protest might also encourage individuals to express views they might otherwise conceal. Presence in a crowd creates the possibility of deindividuation occurring, which might drive individuals to extreme acts that they would otherwise avoid (Douglas, 2010).

These objections may be addressed by emphasising the importance that the shaming of audience vigilantism must be reintegrative. As in the previous example of named-and-shamed KKK members, those identified as white nationalists should be assisted in coming to recognise the prejudices that drive their views and to renounce their earlier racist beliefs. As with reintegrating former KKK members, it will be difficult for many to readmit those who previously supported racist views back into their community. Nonetheless, such acceptance, after those who have been shamed feel genuine guilt for their actions, is necessary to reduce the risk of ostracism, which would further alienate and radicalise those with racist and prejudicial views.

The example of doxing Antifa activists is crucial to establish the importance of the legitimate social need criterion. If the activist’s cause is irrelevant to this evaluation, any moral permissibility granted to the doxing of far-right activists would also hold for doxing Antifa and other social justice activists. Both Antifa and far-right groups attempt to suppress opposing political speech, and both perceive themselves as defending their communities. However, the legitimacy of their perceptions of their community and the threats they are reacting against differ significantly.

Far-right or ‘alt-right’ groups that attempt to suppress subversive expression believe they are defending their community against cultural, moral and/or social corruption. However, their flawed or self-serving understandings of culture and race undermine their conception of ‘community’. The concept of ‘whiteness’ motivating many far-right and alt-right activists incorporates (at least elements of) white supremacy, the ideology that ‘white people’ are in some way intrinsically superior to other peoples, and this superiority justifies these people having unearned privileges and power in society (Beirich & Potok, 2011). The doxing of anti-fascist and other social justice activists by far-right and
alt-right activists is the use of audience vigilantism to silence challenges
to white privilege. To meet the legitimate social need criterion, far-right
and alt-right activists would need other groups in society to accept the
legitimacy of white privilege (and that others must therefore accept
their intrinsic inferiority). Any society that claims to accept political and
social equality (such as liberal democracies) must reject such claims.
Fascists and far-right extremists therefore lack the legitimate social need
criterion for audience vigilantism.

Conclusion

Vigilantism of any form will always be morally controversial. Nonetheless, for those who have legitimate concerns about the ability
of law enforcement to protect themselves and their communities from
harm, vigilantism may be a defensible option. Digital vigilantism, in the
form of doxing those who threaten harm to communities, is a viable
option if it incorporates reintegrative shaming. Without the possibility
that those who are exposed by digital vigilantism can be reintegrated
into their communities, DV risks further alienating them and reinforcing
their extreme views.

Not all justifications for doxing transgressors are equally legitimate.
Hate speech, however, certainly is a legitimate concern. Where the law
prohibits hate speech, its victims should seek legal avenues against it. If
the law is silent about serious expressions of hate, those who face them
may have a legitimate reason to turn to the use of audience vigilantism
(including doxing) to defend themselves, provided that this serves as
the start of an effort to reintegrate proponents of hate speech into a
tolerant society that respects others.

References

Paste Site, https://pastebin.com/wbvP95wg

anonymous-opkkk


