Incels on Reddit: A study in social norms and decentralised moderation
by Rosalie Gillett and Nicolas Suzor

Abstract
This paper examines the challenges of combating misogyny and harmful social norms on the social networking and news site Reddit. Reddit has a long history of hosting racist, misogynistic, homophobic, and otherwise harmful communities. Incels — an Internet subculture that ascribes to deeply misogynistic beliefs — flourished on Reddit before several major subreddits were progressively banned. In this paper, we examine four incel subreddits to understand how the dynamic tensions between Reddit’s site-wide rules and the rules of particular incel subreddits play out in practice. We draw on archived Reddit data to provide a detailed analysis of how incel subreddits adapt to and resist pressure from Reddit and other users to comply with the site-wide rules and develop and maintain their rules and norms. Our study suggests that punitive measures, such as the threat of stricter rules and increased external moderation for explicitly prohibited content and behaviour is insufficient to foster prosocial subreddits. We show how stricter enforcement of the formal site-wide rules did not challenge or displace the underlying ideologies that foster toxic communities. Instead, we find that incels justified their toxic behaviour in the face of heavy external criticism. To cultivate safe and inclusive digital environments, we argue that Reddit must focus on improving the culture of subreddits. We conclude that driving real change when moderation is decentralized requires the committed and supported participation of moderators who can undertake the extensive work of tackling the underlying identity and ideology that brings hateful communities together.

Contents
Introduction
Background
Research methods
Results and discussion
Shared beliefs, defensive rules, and neutralization techniques
Blunt top-down governance and limited decentralised moderation
Similar subreddits, different norms
The problem with incels is incels
Conclusion

Introduction
On 23 April 2018, a man drove a van into a crowd of people, killing 10 and injuring 16 others in Toronto. In a police interview after the attack, the man claimed to have been radicalised online and described himself as an ‘incel’ (short for ‘involuntary celibate’). The term ‘incel’ was originally coined to describe an online group started by a female undergraduate student for lonely people in the late 1990s; it has now come to represent a
In this paper, we argue that ‘content moderation’ is insufficient, both as a practical tool and a conceptual frame, to address the challenge of improving governance of extremist or toxic communities on major social media platforms. Content moderation focuses on the problem of enforcing rules about individual pieces of content at scale (Gillespie, 2018). We use archived Reddit data to explore the cultural norms of extremist and potentially toxic communities that can be influenced by the operators of social media platforms, other users, and external pressure and criticism. Our analysis examines the relationship between rules and cultures — looking specifically to understand what governance measures are most suited to tackling the underlying cultural norms that foster and enable abuse. Drawing on Sykes and Matza’s (1957) neutralization theory, we show how incels worked to resist increasing pressure to comply with Reddit’s site-wide rules.

Reddit is particularly notable among major platforms for its (partially-)decentralised approach to content moderation. Reddit relies heavily on the volunteer moderators of its individual subreddits to set and enforce rules about acceptable behaviour. We argue that Reddit’s initial attempts to convince the moderators of incel subreddits to abide by the site-wide rules did not work to change the misogynistic culture of the main incel subreddits. Incels used their ability to set and enforce rules, combined with Reddit’s automated moderation tools, to try to technically comply without changing their shared misogynistic ideologies. In response to external pressure, the incel communities we studied moved to enforce stronger boundaries that excluded dissenting opinions and meaningful attempts to address misogyny. Driving change in toxic communities requires an ongoing commitment to change cultural norms from influential members of that community. The threat of punitive escalating sanctions based on individual breaches of content policies ultimately led to incel communities being driven off the platform — not significant cultural change. To effectively address toxic cultures, we suggest that Reddit and other platforms must focus on managing the shared social norms and ideologies that bring hateful communities together.

Background

Reddit has been plagued by communities dedicated to misogynistic and other hateful content (Gibson, 2019; Milner, 2013; Topinka, 2018). The social news platform, self-described as “the front page of the Internet”, hosts the voices of over 430 million active users per month in over 130 thousand subreddits (Broach, 2021). While the platform enables identity exploration and connection (van der Nagel and Frith, 2015), Reddit has drawn extensive public criticism for hosting subreddits that advocate white supremacy, disparagement of minority groups (Topinka, 2018), and violence against women (Gillespie, 2018; Massanari, 2017). During #GamerGate, a misogynistic campaign of harassment against women in the gaming industry, Reddit’s “platform and algorithmic politics” provided “a fertile ground for anti-feminist and misogynistic activism” [1]. This fertile ground is where incels gained subscribers and notoriety.

Created in 2013, r/Incels was the first prominent subreddit dedicated to incels. The subreddit marketed itself as a support group for men who were unable to form romantic or sexual relationships; however, using a distinct lexicon that is commonly shared by the broader ‘manosphere’ (Waśniewska, 2020), subscribers’ discussions exhibited the deeply misogynistic brand of antifeminism that they ascribed to (DeCook, 2019; Ging, 2017). r/Incels gained almost 40,000 subscribers before Reddit banned the community on 2 November 2017. Reddit’s move to ban the subreddit came shortly after the platform updated its rules to prohibit content that “encourages, glorifies, incites or calls for violence or physical harm against an individual or group of people” (Papadamou, et al., 2021). The new r/braincels subreddit, along with some external discussion forums, grew rapidly in popularity in the immediate aftermath of the ban.
Incels have been the subject of a great deal of criticism from the media. The extent to which incels have perpetrated acts of physical violence has raised serious concerns around their influence on subscribers’ behaviour (Maxwell, et al., 2020). r/braincels gained increasing media attention in April 2018 after the Toronto van attacker described his affiliation with the community. This act of extreme violence showed that incels’ “war on women” was not “merely metaphorical” (Bratich and Banet-Weiser 2019, 5019). Criticism also came from other Reddit communities; in May 2017, for instance, a Reddit user (or ‘redditor’) created r/IncelTears, a ‘watchdog’ community, which was dedicated to criticising incel behaviour on r/braincels (Dynel, 2020; DeCook, 2019). It was not until r/braincels had grown to over 50,000 members in September 2018 that Reddit quarantined the subreddit, making its content invisible to users who were not registered as members of the subreddit (Papadamou, et al., 2021). Reddit finally banned the community for breaching its site-wide rules prohibiting bullying and harassment one year later in September 2019.

Several other communities dedicated to people who struggle to establish romantic and sexual relationships remain active on Reddit. At the time of writing, r/ForeverAlone, described as a support group for people who struggle to establish romantic connections, has surpassed 170,000 members. While scholars and popular media consider the subreddit to be a part of the broader collection of incel communities (see, for example, Papadamou, et al., 2021), r/ForeverAlone has distanced itself from incels, banning “any incel references, slang, or inference” (r/ForeverAlone, 2020). However, r/ForeverAlone has also faced considerable scrutiny, given that several mass murderers are thought to have been members of the subreddit prior to committing their crimes (Dickson, 2019).

Reddit’s distinctively user-driven content moderation system relies on volunteer moderators (“mods”) to govern their own subreddits (Jhaver, et al., 2019; Matias, 2016). Reddit users are accordingly subject to the platform’s overarching rules, informal norms and policies (‘reddiquette’), and the distinct cultural norms and self-imposed rules of each subreddit in which they participate (Fiesler, et al., 2018). Local social norms shape behaviour and set the expectations of acceptable conduct (Netanel, 2000; Newby, 1993), and differences in how individual subreddits are moderated means that the established behavioural norms and culture of communities can differ greatly (Gibson, 2019). Reddit’s community moderators play a unique governing role, where they “occupy a difficult position between Reddit and their user community: beholden to both, belonging to neither” [2]. Reddit provides the tools that moderators use to regulate, including a bot called ‘AutoModerator’, which subreddit moderators customise to regulate their individual communities (Jhaver, et al., 2019). Users, then, exert a governing influence “through the instantiation of norms and platform practices” [3].

Unlike other large digital platforms, Reddit has taken a largely hands-off approach to content moderation (Massanari, 2020). In some circumstances, encouraging smaller groups to make rules that fit their needs might be preferable to a set of one-size-fits-all rules that apply across an entire platform. The early cyber-libertarian literature emphasises the benefits of decentralisation: where users are able to self-select into communities whose norms align with their values and are closely tailored to the needs of the local group, that set of norms is likely to work better, for participants, than more general rules imposed from outside (Johnson and Post, 1996). Community moderators, who understand the culture and are known and respected by other members, may also be better placed to enforce rules than the armies of anonymous human moderators and machine learning classifiers that major platforms deploy to moderate content (Roberts, 2019). But communities have externalities; the development of a toxic culture on a given subreddit, where hatred is allowed to fester, spreads to affect others.

Reddit’s enormous userbase and the extent to which the platform provides a fertile ground for toxic communities to develop and grow (Topinka, 2018) has the potential to expose many to — and normalise — deeply misogynistic attitudes and beliefs. Without intervention and strong governance, hateful norms can become entrenched and continuously reinforced by existing and new participants (Rajadesingan, et al., 2020). Incel culture is worrying because ordinary, everyday acts of abuse and hate are the bedrock for an ideology and culture that supports violence against women (Kelly, 1987). Because of the visibility of potentially harmful content on Reddit, there is a real risk that this serves to legitimise and reinforce hateful messages (Topinka, 2018; Byerly, 2020).

Reddit takes action against subreddits when they become too controversial. On Reddit, quarantining can slow down the growth rates of harmful subreddits; a recent study shows that quarantined subreddits struggled to attract new users (Chandrasekharan, et al., 2022). Banning entire communities may be somewhat effective to
Incels on Reddit: A study in social norms and decentralised moderation

close down their presence on the platform — although the extent to which deplatforming detoxifies the Internet more broadly is not clear (Rogers, 2020). It is likely that deplatforming pushes groups to unindexed corners of the Internet, which are “far more resilient and harder to moderate, allowing otherwise radicalizing, false, or hate-filled content to persist” (Jardine, 2019). However, deplatforming from major sites is likely to shrink harmful communities to a more highly motivated core of users; less-engaged users will discontinue participating in these spaces (Chandrasekharan, et al., 2017).

Reddit’s power to influence its subreddits is deliberately limited by their decentralised architecture. Reddit is heavily reliant on its volunteer moderators to take responsibility for their subreddits. This moderation system is inextricably linked to the culture of the subreddit; effective rules have to be consistently enforced and align with community expectations (Ivaturi and Chua, 2019), and the moderators themselves are often core members of the subreddits they oversee (Seering, et al., 2019). In this paper, we make use of Sykes and Matza’s (1957) neutralization theory to understand how incels reacted to external pressure from Reddit and other redditors to change and enforce rules to discourage misogyny. This criminological approach highlights how people distance themselves from taking responsibility for their actions (Sykes and Matza, 1957). The theory has been widely used to examine the techniques that people rely upon to downplay their harmful behaviour and resist allegations that their actions are wrongful (Hwang, et al., 2016). We begin from the assumption that mods may not foster prosocial communities if they are unwilling to recognise that their behaviour is harmful or take responsibility for the cultures that they help to create.

Reddit’s userbase is deeply conflicted between a desire for subreddit autonomy and demands for greater central oversight. In its early years, Reddit’s prioritisation of freedom of speech, subreddit autonomy, and user voting meant that its site-wide rules were little more than the minimum required to comply with relevant laws and protect the platform’s brand and security. In the midst of a major controversy over the abusive leaks of intimate photos of celebrities and other women, former Reddit CEO Yishan Wong explained that while Reddit ‘sympathized’ with the victims for the harm caused by misuse of the site, it would generally not act to address ‘morally objectionable’ behaviour, on the basis that ‘each individual is responsible for his or her moral actions’ (Wong, 2014). In more recent years, in the face of steadily increasing pressure from the media and from its own users, Reddit has taken steps to more extensively regulate the content on its service (Gillespie, 2018). In 2020, the Black Lives Matter movement prompted an r/AgainstHateSubreddits moderator to write an open letter calling on the company to enforce its rules against hate speech and ban hateful communities. The letter was co-signed by moderators of over 800 communities representing hundreds of millions of redditors (r/AgainstHateSubreddits, 2020). Within days, Reddit banned several subreddits including the 700,000 strong r/The_Donald — a community that was dedicated to Donald Trump’s supporters.

There is little consensus of the extent of the responsibilities of platforms for the behaviour of their users. Initially, platforms like Reddit worked hard to distance themselves from taking on responsibility for what their users did, supported by U.S. law that immunises them from many legal threats arising from user-generated content (Gillespie, 2010). Platforms, however, are never neutral. Their rules and enforcement systems, technical design and affordances, their algorithmic curation and recommendation systems, their marketing and internal messaging all work to shape the cultural norms that govern what conduct and content is acceptable, encouraged, and rewarded. Governments and private actors are now increasingly expecting platforms take a more active role in governing their users’ behaviour. In this context, we set out to examine the relationship between Reddit and its subreddits at a time of growing controversy over the acceptability of incel culture on the platform. We examine how incel communities reacted to rising external criticism and the threat of punishment from Reddit — and what impact this had on the development of their rules and norms. The ongoing complaints from the community serve to highlight the fraught and heavy investment that Reddit’s decentralised model requires from its volunteer moderators.

**Research methods**

We approached this study using a range of different methods that helped to illuminate the development and contestation of social norms and formal rules. To examine the external pressure on Reddit and in incel
Incels on Reddit: A study in social norms and decentralised moderation

We initially extracted 2,033 articles from the Factiva news database, using the search term ‘incel*’. We read a small sample of randomised articles, and then read deeper into the major events and controversies that generated many news and opinion stories. Ultimately, this analysis was helpful to ground our understanding of media pressure, but was also limited in that the bulk of articles were published after incels became highly topical in 2017 and later.

Next, we extracted archived Reddit data to examine user discussion in r/Incels, r/braincels, r/ForeverAlone, and r/IncelsWithoutHate. We used the social media data collection, analysis, and archiving platform Pushshift.io, which makes Reddit data available to researchers (Baumgartner, et al., 2020). The Pushshift datasets are not complete, but they provide a highly useful archive of Reddit posts and comments that includes material from the now deleted r/Incels, r/braincels, and r/IncelsWithoutHate subreddits. We initially collected 13 million comments from Reddit between 2011 and the end of August 2019 that mentioned the term ‘incel’, spread across many different subreddits that we categorised into incel groups, critical subreddits, and generally unrelated subreddits. To help us analyse the large numbers of comments we were interested in, we reproduced posts and comments in a threaded format that is similar to how comments are displayed on the live Reddit site. The dataset generally does not include the text of comments that are subsequently deleted, but it is possible to see that comments in a thread have been deleted.

We located discussion about Reddit and the individual communities’ rules by focusing our analysis on keywords related to moderation and discussion including and surrounding the subreddits’ moderators. We used a list of keywords that we developed iteratively that could most accurately return explicit discussion about governance and exclude the much larger quantity of general shitposting (Daly and Reed, 2022) that quickly became familiar and repetitive (for our purposes). We went through several iterations of reading random samples of posts and comments in order to develop a set of keywords that would allow us to find explicit discussion about moderation and governance of incel subreddits. We found the most active discussion in three large subreddits: the active incel subreddit r/ForeverAlone, and the now banned incel communities r/Incels and r/braincels. We also included a much smaller subreddit, r/IncelsWithoutHate, a subreddit that was created directly after r/braincels was banned whose mods explicitly sought to create an incel community that was compliant with Reddit’s site-wide policies. Focusing on these subreddits, we identified their historical moderators using the Wayback Machine — a digital archive of portions of the World Wide Web. We also used the Wayback Machine to develop a log of the subreddit descriptions and rules that guided these communities, focusing on how these statements changed over time.

Ultimately, we assembled chronological datasets of 500 threaded posts and comments from each of the three subreddits. These samples include posts and comments by moderators, as well as posts that mention moderators by name or explicitly used a set of moderation-related keywords. For all matching comments and posts, we collected every available comment up and down the thread of replies to each of these posts or comments. We discarded comments that were not ancestors of or replies to the comments we identified as potentially relevant, leaving nearly 20,000 comments across the three subreddits. Given that users frequently trigger Reddit’s AutoModerator tool, which replies with very similar pre-programmed responses, we excluded these comments from the samples. We did, however, analyse the AutoModerator posts in a separate sample, since this feature operates as an important governing tool in individual communities.

We undertook a qualitative analysis of the subreddits’ descriptions, rules, posts, and comments. Guided by neutralization theory (Sykes and Matza, 1957), we used open and axial coding techniques to comprehensively analyse the data, looking for explicit discussion about the criticisms levelled against incel subs. We paid particular attention to the responses of incels to criticisms about the rules and norms — what the communities allowed, what they prohibited, and the approach the mods took to justifying, changing, and enforcing the rules. Our analysis shows how incels engaged with criticism and resisted external pressure to change their norms, and how and when the moderators enforced the rules. We chronologically read posts and comments in each subreddit sample, ensuring that we read posts from each month in order to understand the evolution of moderation discourse. We read many thousands of comments and posts until we reached saturation. Reading the discourse around moderation enabled us to focus on the explicit discussion of rules, enforcement, and the evolution of social norms about permissible content and behaviour in each subreddit.

Following Jane in her work on online misogyny, we have included direct unedited quotes from the incels
Incels on Reddit: A study in social norms and decentralised moderation

Results and discussion

Misogyny is deeply embedded in the culture of the incel subreddits. Some of the most influential incels promoted physical and sexual violence, as others have also found (Scaptura and Boyle, 2020). Incel communities have developed a vocabulary that is shared among adjacent communities in the manosphere (Wasniewska, 2020; Hoffman, et al., 2020). This vocabulary serves to ‘other’ and sexually objectify women; it includes the dehumanising term ‘femoids’ or ‘foids’; derogatory stereotypes of ‘stacys’ and ‘beckys’; ‘feminazis’; or the slut-shaming term ‘roasties’ (Glace, et al., 2021). They also distanced themselves from other men, referring to conventionally attractive men as ‘chads’, and people who do not identify as incels are ‘normies’ or ‘normans.’ This shared language enabled incels to build and preserve a strong group identity (DeCook, 2019) that they would defend strongly against incursion.

Incels came under a great deal of social pressure to improve their cultural norms as they became more visible in the mainstream media. The Toronto van attack illustrated the extent to which harmful online subcultures may play a role in influencing individuals to perpetrate acts of serious physical violence. In the aftermath of the attack, several Reddit communities began to pay close attention to incels’ behaviour. Subscribers to the popular incel-critical subreddit r/Incels Tears, for instance, commonly browsed the incel communities, screen captured their posts, and published them in their own subreddit to scrutinise their behaviour (Sang and Stanton, 2020). Later, in 2019, r/braincels received a great deal of attention from the media and redditors surrounding the release of Joker, a movie about Batman’s villainous nemesis. The concern was caused by a man who, in 2012, entered a United States cinema where he shot and killed 12 people and injured 70 others at a screening of the Batman movie The Dark Knight Rises. Given the 2019 Joker film’s sympathy towards violent men, influential film critic David Ehrlich argued that the movie is “a toxic rallying cry for self-pitying incels.” Incels themselves discussed the possibility of gun violence surrounding the film, with one questioning whether it would “inspire people to go ER” (cited in Owen, 2019) — a reference to an incel who murdered six people in California in 2014.

Shared beliefs, defensive rules, and neutralization techniques

As the growing incel communities became subject to increased scrutiny and were challenged for their beliefs, they developed a set of defences against outsiders. Reddit’s community-led moderation provided a petri dish for incel communities to grow rapidly with minimal external interference. The two major incel subreddits, r/Incels and the later r/braincels, created and maintained a set of formal rules and cultural norms that promoted and protected their distinctive misogynistic ideologies.

Incels frequently referenced evolutionary or pseudo-scientific theories to provide intellectual legitimacy for their shared beliefs. These beliefs were upheld as truths that came to define the community, as a r/braincels participant (2018) illustrated: “Women can’t be virgin or incel. They can have sex all they want but when no Chad wants to fuck them anymore they say ‘wHeRe DiD aLl ThE gOoD mEn Go?!!!!!!!7.’” The incel discourse strongly reflects the ‘red pill’ language and themes that are familiar across many other facets of the manosphere (Jones, et al., 2020; Van Valkenburgh, 2021). We commonly saw, for example, that complaints about misogyny were dismissed as ‘uncomfortable truths’, as one r/braincels participant (2018) said when another commenter suggested that they be banned for slut-shaming women: “Lol, banning for what, for stating the truth? Fuck off.”

Sykes and Matza’s (1957) neutralization theory can help us understand how incels justified and devolved culpability for their harmful behaviour. Their shared beliefs in “turbocharged genetic determinism” [6] enabled
Incels on Reddit: A study in social norms and decentralised moderation

Incels to deny responsibility for their lack of romantic relationships and provided an opportunity to justify their misogynistic attitudes (Baele, et al., 2021). According to one r/Incels participant (2017), a combination of “social conditioning, bad genetics and oppression from women” prevents the community from forming romantic relationships. Incels used this narrative to claim that they were “helplessly propelled into new situations” whereby factors out of their control dictated their and others’ behaviours.

The shared beliefs held by the community were enshrined in the rules and social norms of the subreddits and were strongly defended. These rules were deployed by r/Incels and r/braincels moderators to prohibit discussion that challenged incels’ misogynistic beliefs: “Accusatory posts that generalize the entire subreddit (e.g., ‘Why do you creepy virgins hate women so much?!’ or ‘Are you people for real? Is this a joke?’) will be treated as spam and removed” (r/braincels moderator, 2018). These explicit rules to shield incels from criticism — what they saw as ‘inflammatory’ and ‘accusatory’ comments — were enforced primarily against redditors who challenged their beliefs. These rules evolved and became more specific over time to address perceived threats to the essential beliefs that united the community. In September 2016, for example, r/Incels adopted a rule that prohibited discussion that called into question their fundamental assertion that women do not generally struggle to attract sexual partners:

Those claiming there are as many female incels in the same situation as male incels are banned. Most can agree that women can be incel in some situations, but saying that there are many incel women in the same situation as incel males will get a warning and your comment removed.

A key part of neutralization is the work of an offender to shift “the focus of attention from his own deviant acts to the motives and behaviour of those who disapprove of his violations” (8). We saw much of this work in the policing of boundaries and ongoing efforts to remove critics and dissenters from each subreddit. It was common to see incels promote their shared ideology by condemning and excluding their critics: “This subreddit should be a male only sub, specifically for Incels. Letting bitches post here should not be tolerated in whatever manner. They bring nothing to the table except for useless advice, meaningless platitudes, and overused sayings such as, ‘this is y ur incel!!!’. [...] I’ve personally seen a few cum dumpsters comment here regularly, and I wonder to myself why these cunt[s] aren’t fucking banned” (r/Incels participant, 2017). It was common to see incels respond to criticism in this way, attacking or seeking to exclude people who challenged their beliefs: “I don’t know why we entertain the idea of females organisms posting on this sub. They don’t say anything of value at all. Just the same old crap like, ‘this is why you’re incel’, ‘personality is what matters’, ‘treat women with respect’, and blah blah blah” (r/Incels participant, 2017).

Incels cast themselves as victims, and justified these exclusionary policies by claiming that the subreddits were “support group[s]” for bullied and oppressed males (r/Incels participant, 2017). The irony here, apparently lost on incels, is that they used this framing in an effort to neutralize their vitriol and hostility toward outsiders (Maxwell, et al., 2020): “most incels are not what normies claim them to be. On the other hand, most normies on this sub are exactly the subhuman scum that they’re generalized as. [...] The only difference is that it’s a sub for incels, so harassing incels here is against the rules, but doing the same to normies who come here and act like scum is allowed” (r/Incels moderator, 2016).

Because subscribers to both r/Incels and r/braincels argued that ‘normies’ harassed and bullied them, they used this rhetoric to absolve incels of responsibility for their harmful retaliations. One r/Incels participant (2016) who, within weeks, became a moderator of the subreddit, explained that the mass murderer Elliot Rodger’s actions were justified because of how ‘normies’ have treated incels:

[normies] are scum, they only come here to bully, harrass and taunt incels. AND they are very, very stupid. If they think harassing, and hating on incels is going to make us like women or normies, just how retarded are they? It just proves that us incels, even if we aren’t politically correct, are by far morally and intellectually superior to the normies. It also justifies the actions of incel men such as Elliot Rodgers. That’s just a fact, because all the
Incels on Reddit: A study in social norms and decentralised moderation

people tormenting and hating on incels deserve to be treated like enemies, they are arguably as bad as terrorists.

Incels’ condemnation of women and non-incels as ‘scum’, ‘bullies’, and ‘terrorists’, “incorporated the language and forms of warfare, revolt, and terrorism ...” [9] and played an important role in discourse that rejected their own responsibility for violence and harmful behaviour. To these incels, the community was at war with anyone who challenged their beliefs. Positioning their condemners in this way enabled incels to frame themselves as victims who needed to defend themselves from harassing outsiders — an ‘othering’ tactic, which creates an ‘us’ versus ‘them’ mentality, as observed in other subreddits including The_Donald (Gaudette, et al., 2021) — and perpetuate the myth of male victimhood at the hands of women (Marwick and Caplan, 2018; Banet-Weiser, 2018). As a result, they could position “themselves as noble rebels against a common cause” [10], justifying harassment towards critics — and worse — as a defence mechanism (Marwick and Caplan, 2018).

Incels also dealt with criticism by blaming their critics for the worst behaviour in their subs. Incels argued that their subreddits were being ‘brigaded’ by other subreddits, and that many of the most inflammatory comments in their communities were ‘false flags’ posted by critics disguised as incels in order to prompt action from Reddit. It is difficult to prove or disprove these assertions, and the potential of false flags gave incels a mechanism to disengage or delegitimise critiques that would otherwise be very difficult to defend.

As the incel subreddits became more defensive, they more rigorously policed their boundaries. Given the increasing profile of incel subreddits, both r/Incels and r/braincels made use of Reddit’s automated tools to help with the ongoing work of keeping outsiders out. Reddit’s ‘AutoModerator’ tool (‘Automod’) enables the moderators of individual subreddits to create and enforce rules, including removing or responding to comments that use particular keywords. The incel communities relied on Automod to automatically delete critical comments, ban users, and to reinforce shared norms when they were questioned.

One particularly striking example of their defensive strategy is how incel subreddits configured Automod to respond to comments that criticised incels — including what one user described as “even the mildest suggestions” for self-improvement and self-help. Automod was used to remove and respond to keywords associated with counterspeech that could have been useful in helping incels acknowledge their responsibility for their misogynistic beliefs and behaviour. The r/Incel moderators customised Automod to post lengthy written responses to common criticisms levelled against them by outsiders. For example, in response to outsiders who called incels out for being ‘entitled’ (to sexual activity and women’s attention), the Automod responded with:

‘Entitled’ is a fairly meaningless word that does not accurately describe most of the incels on this sub or elsewhere on the Internet. On one extreme, the word ‘entitled’ is used to describe men how literally believe that they should be able to force women to date them. On the other extreme, the word ‘entitled’ is used to describe men who are merely frustrated or sad that they can’t find anyone at all to date them. This type of frustration is reasonable, since sex and romantic relationships are regarded by many as one of the most fulfilling things in life. The problem with using the word ‘entitled’ for both categories of men is that it lumps them together, demonizing men in the latter category by comparing them to men in the former category. If you wish to criticize our views, please be more specific than merely calling us ‘entitled.’

Automod was key in helping the incel subreddits to neutralize criticism. With its assistance, the moderators of the incel subreddits worked consistently to mitigate their subreddits’ exposure to insults and criticism. The absence of diverse points of view worked to validate (Gaudette, et al., 2021) and normalise (Jones, et al., 2020) extreme views.

The prevalence of template responses and the extensive removal of counterspeech shaped the discourse in a way that helped incels legitimise their position. Reddit’s automated tools are designed to help communities cohere around shared ideas. Their use by incel subreddits might not have been specifically anticipated by Reddit, but the
tools were used by incels in the way that they were intended to work. This type of predictable repurposing of its ostensibly neutral tools is an important illustration of how Reddit’s approach to moderation has worked to support the development of toxic communities. As Nakamura (2013) points out, the use of generic features of Internet platforms to spread hate is ordinary — not a ‘glitch’ or aberrant failure of moderation, but a predictable and apparently inevitable consequence of public online communications technologies. In distancing themselves from the more extreme behaviour they understood as harmful, incels resisted pressure for deeper cultural change. It appears that incels, like Reddit, might be understood as viewing incel violence and the most extreme transgressions of site-wide rules as aberrations, rather than predictable manifestations of cultural and ideological norms.

Blunt top-down governance and limited decentralised moderation

The increasing toxicity and size of the incel subreddits placed them at increasing conflict with Reddit’s site-wide rules. When Reddit takes action against subreddits with toxic cultures, its interventions are punitive and blunt. The platform quarantines toxic communities — a move which makes them less visible to non-subscribers — or bans entire subreddits that violate the site-wide rules. As the incel subreddits received more attention and pressure from outsiders, Reddit threatened to take more serious action against subreddits that continued to break the platform’s site-wide rules. Ultimately, Reddit’s threats to quarantine or ban communities did not carry enough weight to meaningfully shape user behaviour and norms in the incel subreddits (Copland, 2020).

Reddit’s decentralised governance model means that toxic communities are given more room to fester. As a platform that devolves moderation responsibilities to volunteer moderators, Reddit’s mode of governance enables users to cluster in groups where the norms are much more closely aligned to their own values and interests. Moderation teams play an important norm setting role in their communities (Park, et al., 2016; Fiesler, et al., 2018). This is part of what makes Reddit a platform where niche or marginalised groups can flourish and have meaningful online experiences (van der Nagel and Frith, 2015).

Generally speaking, both r/Incels and r/braincels recruited moderators who self-described as incels. On both r/Incels and r/braincels, the majority of moderators were clearly not interested in challenging the misogynistic values of their subreddits. Moderators are frequently drawn from the most active members of the community; several moderators actively legitimised hate speech and expressed pride in the communities’ reluctance to remove harmful content.

After r/braincels was quarantined, participants increasingly expressed concern that Reddit might take further action and ban the subreddit. The moderation team appointed more moderators — to little tangible effect. Some of the discourse recognised that moderators, who are generally highly active members of the community, are likely to reinforce the subreddit’s existing ideology: “Appointing more Incels as mods isn’t fulfilling the quarantine requirement specifically. The charge levied was ‘excessive misogyny’. ... misogyny needs to be minimized if not eliminated” (r/braincels moderator, 2018). Others noted that the threat of quarantine actually works to strengthen the communities’ existing propensity to keep outsiders out: “Quarantined subs are literally the best subs” (r/braincels participant, 2018), to which another r/braincels participant (2018) agreed: “Keeps the Normans out.”

At r/braincels’ inception, after Reddit banned r/Incels, the new subreddit’s rules prominently asserted compliance with site-wide rules. There was a clear threat that Reddit would ban the new subreddit if it continued to be hateful:

You should be fucking happy, idiots, you do realize if you turn this sub into /r/incels 2.0 it’ll be banned by February, right? And not because it will be a collection of ugly men telling the truth or whatever bullshit rationalization, but because it’ll fall into the same madness and depravity of the previous sub. You need someone to moderate you, someone that won’t just enforce the
Incels on Reddit: A study in social norms and decentralised moderation

echo chamber you so desperately want. For fuck’s sake, what do you even gain out of everyone patting you on the back and telling you obvious bullshit is true? (r/braincels participant, 2018)

This suggestion, however, was controversial. A r/braincels moderator (2018) was among those who disagreed: “there’s no excuse for turning this sub into another trap like r/foreveralone.” The explicit adoption of hateful ideologies is a core part of what binds incel communities together. One r/Incels participant (2016) reflected approvingly on the subreddit’s moderation approach compared to other subreddits: “The blasé attitude of the mod team towards rape, murder and pedophilia are what sets this sub apart”. r/braincels subscribers fought to maintain the distinct ideology of incels; subreddits like r/ForeverAlone, which place some of the responsibility for men’s loneliness on the individual men, threaten that ideology by limiting their ability to neutralize their responsibility.

There were some moderators of r/Incels and r/braincels that worked to change the culture. Incels sometimes commented that their moderation teams were not misogynist, pointing to the appointment of a small number of moderators who identified as women (but not incels). These women were among a small number of moderators who seemed to be most committed to removing abusive comments and limiting the most extreme discussion within the incel subreddits. One of these women explained that she volunteered for a moderation role after learning that a family member was affiliated with incels and that she worked hard to try to help people she saw as disillusioned or upset but potentially open to changing their attitudes.

This small group of moderators appeared to undertake an outsized proportion of the work of moderation, pushing the subreddits to more closely align with Reddit’s site-wide rules. This work was almost always framed as an effort to mitigate the risk of enforcement action from Reddit against the community:

although posts about murder fantasies or wanting to kill someone aren't explicitly against the sub’s rules, it is possibly against reddit’s rules so to avoid too much negative attention and risk of quarantine these will likely be removed. Posting about how when you see a couple holding hands you want to “fucking bash their heads on a wall and repeatedly stab them with a rusty knife until they can’t feel themselves anymore” doesn’t do much good for the sub. It doesn’t add value to the subreddit and potentially violates reddit’s rules, so to be safe, things like fantasizing about murdering a couple will be removed. (r/Incels moderator, 2016)

Despite the extensive labour these women donated to enforce the rules of each subreddit, their efforts were frequently met with extreme hostility and immense resistance from the community who consistently degraded them. The moderators who identified as women faced frequent abuse from participants — and sometimes even other moderators — including one who was doxed severely and consequently stepped down from the role. As the threat of further action by Reddit loomed, participants became more tolerant of stronger moderation, although they were still often explicitly hostile to outsider moderators.

In general, most members of the moderation teams seemed reluctant to enforce Reddit’s site-wide rules, effectively undermining the work of those mods who sought to improve the community’s social norms. Even where they did enforce the rules, many moderators knew that a low degree of moderation of content would not threaten the subreddits’ shared ideology, as one r/Incels moderator (2016) explained: “even if we deleted one post, there are still lots of other posts with extreme viewpoints from the last few days. So this sub is in no danger of becoming similar to ForeverAlone.”

Reddit seems to have few effective tools to influence subreddits that are developing a toxic culture; quarantining or banning are exceptional events. The experience of the incel subreddits shows that the threat of quarantine or banning was not sufficient to materially alter their behaviour or tackle their harmful social norms. When the Incel subreddit was finally banned, it was quickly replaced by r/braincels, which kept the same distinctive misogynistic vocabulary and quickly outgrew r/Incels in posts and subscribers. In terms of distinct accounts, the userbase of the subreddits are quite different. It is possible that many incel participants created new accounts
after r/Incels were banned, of the nearly 50,000 distinct accounts that made more than five comments or posts in r/Incels and r/ForeverAlone, under 4000 made more than five comments or posts in r/braincels. But the culture remained very similar, and the new moderation teams of r/braincels did not really seek to actively counter misogynistic discourse. Indeed, to the extent that r/braincels moderators responded to Reddit’s threats, they seemed to be mainly concerned with limiting controversy by keeping outsiders out.

---

**Similar subreddits, different norms**

At the same time that the incel subreddits were active on Reddit, similar communities dedicated to those who struggle to establish romantic relationships grew in popularity. Among these subreddits, r/ForeverAlone continues to be a popular “place where people who have been alone most of their lives [can] talk about their issues” (r/ForeverAlone, 2020). The subreddit deals with most of the themes affiliated with incels, but its discourse is quite different. Qualitatively, there is a real sense that participants in r/ForeverAlone are more reflective and committed to self-improvement. Like the incels subreddits, r/ForeverAlone still provides an outlet for people to vent their frustrations about their loneliness.

From its inception, r/ForeverAlone’s moderators positioned the subreddit as an inclusive community. Soon after the subreddit was formed in October 2011, the moderation team created and enforced rules that prohibited abusive behaviour:

**Rules:** Very simple and if they can not be followed the consequences will never be the same!

This is a community where **ANYONE** is welcome regardless of their status so please do not tell people to leave because they are in a relationship or have been in one some of these people want to just give a helping hand.

Trolling will not be tolerated PERIOD! If you see a post that you feel is trolling please bring it to the attention of the mods that is what we are here for.

Have respect for other people. Please this is very key to the community especially since we get some people merely looking for advice and/or need to vent so please welcome them with open arms.

And if you are having a problem with a fellow member please again bring it up to the mods. Do not try fighting back with them unless you can do it in a reasonable manner. (emphases in original)

As the subreddit started to grow, moderators (2012) worked to head off the ideological attacks that became familiar in the incel subreddits:

Do not post anything negative towards other genders, races, and sexual orientations (generalizations, social theories, rants, etc.). Personal experience is not some blanket rule.

r/ForeverAlone continuously struggled with misogynistic behaviour — and struggled to distinguish itself from the incel subreddits. Incels saw r/ForeverAlone as a reference point for the hostile, ‘heavily censored’, misandric environment that the incels subreddits could become if they lost the distinct incel culture. At the same time, r/ForeverAlone subscribers worried that their subreddit might become more misogynistic. After r/braincels was
incels on Reddit: A study in social norms and decentralised moderation

quarantined in 2018, one r/ForeverAlone participant (2018) worried that the subreddit had begun to “toe the line of inceldom”. Several others observed a shift in the subreddit’s culture during the time that they were members: “I guess now’s the time to say it. I: I have been worried for a while. Recently I’ve seen quite a few posts and comments which I didn’t feel were appropriate, and which I felt had hints of incel ideology” (r/ForeverAlone participant, 2018).

On the whole, r/ForeverAlone was mostly successful at maintaining a less misogynistic culture than the incel subreddits. Despite the similarities in topics of discussion, the key difference between the incel subreddits and r/ForeverAlone we found is in how the moderators worked. In the incel subreddits, misogynistic ideology was explicitly and continuously supported. The rules of the incel subreddits were enforced primarily tokenistically, if at all; moderators were active participants in perpetuating the community’s misogynistic ideology. In r/ForeverAlone, misogyny was much more frequently countered by participants and by moderators who challenged the core incel beliefs that women are primarily responsible for their inability to date and find romantic or sexual partners. r/ForeverAlone was notably much more inclusive of women, and as a result seemed much more respectful overall. We suspect that the exclusion of women and women’s perspectives was a core technique in neutralizing incels’ responsibility for failing to establish romantic and sexual relationships — perhaps going some way to explaining just how deeply misogyny was baked into the culture.

The problem with incels is incels

It is difficult to reform toxic subreddits. The core tenets of incel ideology are that society is fundamentally hierarchised along sex and attractiveness lines (Baele, et al., 2021) and that women are intrinsically superficial, since they ostensibly decide who to establish romantic relationships with based solely on physical attractiveness (Hoffman, et al., 2020). Maintaining their beliefs required incels to avoid taking responsibility for their personal challenges by excluding other perspectives. It is difficult to imagine an incel community that is not misogynistic; changing that culture requires much more than merely changing the rules.

A comparatively much smaller subreddit called r/IncelsWithoutHate provides an interesting illustration of the difficulty of separating out misogynistic ideology from the incel culture. At some point, some incels felt uncomfortable with the dominant incel communities and tried to change the culture, albeit with little success. In the months before the first major incel subreddit was banned (r/Incels), a Redditor created ‘r/IncelsWithoutHate’ in an attempt to reform Reddit’s incel communities. To our knowledge, it was the only incel subreddit that actively sought to distance itself from the hateful culture exhibited by both r/Incels and r/braincels. At r/IncelsWithoutHate’s inception, the founder framed the subreddit as a space for users who identify as incels; however, do not feel the same hatred towards women. It is clear from the founding user’s first posts that they attempted to build a respectful incel subreddit:

Welcome to r/IncelsWithoutHate. I wanted to make a subreddit for all the people on incels who don’t feel the same hatred that is often expressed there, or maybe want to stop feeling that hatred. I don’t entirely know what this subreddit will be like yet and I’m very open to suggestions, so feel free to post whatever you think might be helpful. Also it would probably be helpful to have some moderators, so if anyone is interested in becoming a moderator here I would be very grateful if you could pm me. Thanks for reading. (r/IncelsWithoutHate moderator, 2017)

Similar to r/ForeverAlone’s rules, the r/IncelsWithoutHate rules at this time also stipulated that “no hate” was permitted and that “everyone is welcome but in return we expect you to be accepting of each other. No bullying and always follow reddiquette [Reddit’s informal policies]” (r/IncelsWithoutHate, 2017). These rules departed from the dominant incel subreddits’ rules that aimed to prevent outsiders from participating in the communities.

Reddit’s threats pushed r/IncelsWithoutHate to package their ideas in more palatable ways that minimise
external criticism and avoid triggering Reddit’s content moderation systems. Rather than creating a respectful and supportive community, however, r/IncelsWithoutHate continued to normalise incel subculture. The overt conversations about physically harming women that were once commonplace in the banned incel subreddits were replaced in r/IncelsWithoutHate by comments that attribute blame to women for their problems: “Incels Don’t Need Women to ‘Join’ The Community. We need them to Resolve the issue” (r/IncelsWithoutHate participant, 2017). Like the discussions that positioned outsiders as harassers, this narrative reflects how several r/IncelsWithoutHate subscribers condemned women, albeit in less recognisably toxic ways. This framing enabled these subscribers to normalise misogynistic narratives that place women at fault for “the ‘injuries’ dealt to masculinity” \[11\]. And because of this, they continued to repress the wrongfulness of their own behaviour \[12\]. The insidiousness of normalising misogyny should not be underestimated: there is a risk that packaging misogyny in more palatable ways persuades outsiders who are curious about the ideology to join (Vysotsky and McCarthy, 2017).

The extent to which the community departed from its deeply misogynistic predecessors at all was disputable (Kesvani, 2019). While the moderation team focused on more overt manifestations of hate, subscribers remained fixated with their physical appearance and women’s perceived responsibility for their loneliness. Some of the most active subscribers to the subreddit (proudly) argued that incel subculture is built on values that are incongruous with non-toxic behaviour. One r/IncelsWithoutHate participant (2017) reinforced the “misogynistic core” \[13\] that underpins incel culture:

The entire point of incels is hatred. Doesn’t matter what kind: self-hatred, hatred of normies, hatred of women, hatred of Chad. Without that hate we’re just like those whiny pieces of shit over at /r/foreveralone. Self-pity without hatred is disgusting.

As Squirrell (cited in Kesvani, 2019) argues, subscribers to this subreddit “believe the same things as the incels who have been banned from Reddit,” and that “the core of their belief, philosophy, whatever you want to call it, is the same.” Subscribers were overwhelmingly reluctant to change their attitudes towards women; excluding others and hating women is a large part of what it means to be an incel. As Bratich and Banet-Weiser \[14\] put it: “this nihilistic network does not seek to overcome its condition: [incels] accept their fate and even dwell in its finality.” Without incels’ desire to improve, it is not surprising that r/IncelsWithoutHate soon resembled its predecessors. Reddit eventually banned the community for violating the platform’s rules against promoting hate in March 2021.

---

**Conclusion**

Encouraging communities to govern themselves can promote diverse and flourishing subreddits that come together around shared ideas and beliefs (van der Nagel and Frith, 2015). Decentralisation, however, also provides the environment for hateful, insular communities to fester. Reddit’s intervention to quarantine or deplatform entire communities, like action by other social media platforms, usually comes at a time when platforms are under a large amount of public pressure (Ananny and Gillespie, 2016). This is generally too late to make real change — communities have already formed around harmful ideologies.

Promoting positive, prosocial communities requires fundamental changes in culture. This in turn requires (at least) the active participation and ongoing commitment of the leading figures in the community. To the extent that incel moderation teams responded to the threat of Reddit quarantining or banning their subs, they mainly discussed the need to address the most explicit content that violated Reddit’s site-wide rules. But mostly, moderators maintained the incel ideology that binds the community together.

Faced with criticism for their misogynistic beliefs that blamed women for their difficulties developing sexual and romantic relationships, incels responded by turning inwards. Most of the work of incel moderation teams was focused on keeping outsiders out in order to insulate the community from external perspectives that could challenge their beliefs. They legitimised their claims by prohibiting overt criticism and any suggestions that they...
were at least partially responsible for their problems. Incels used various neutralization techniques to position themselves as the oppressed population, rejected the claims of victims by turning to pseudoscientific theories, denied any link to mass murders and other violent acts, and dismissed the most vitriolic statements as ‘false flags’ planted by outsiders.

Perhaps the conclusion here is that the problem with incels is incels; the group cannot be separated from the toxic ideology that forms their shared identity. But the experience of r/ForeverAlone moderation teams in actively resisting this ideology should offer some hope. Ultimately, we suspect that large-scale content moderation is not the best way for major platforms to think about the challenges of promoting good community governance. Contemporary content moderation debates tend to focus on the extremes of overtly hateful content. We suggest that more attention is required (from platforms and scholars) on the ordinary, everyday discourse that enables hatred and hateful behaviour. Changing these deeply held beliefs will require approaches that most likely do not look like the current tools of content moderation. There are clear limits to the impact of measures that focus on individual pieces of content and punishing individual users; one of the critical tasks ahead for platforms and scholars is to explore what works to change the underlying ideology that brings hateful communities together. This raises many unresolved questions for future work: to what extent can platforms reform hateful communities and rehabilitate those who cause harm? What alternative approaches to governance might be effective at scale? Can platforms reform hateful communities, and when is deplatforming the most appropriate response? 

About the authors

Dr. Rosalie Gillett is a Postdoctoral Research Fellow in the Australian Research Council Centre of Excellence for Automated Decision-Making and Society at the Queensland University of Technology.
E-mail: rosalie [dot] gillett [at] qut [dot] edu [dot] au

Professor Nicolas Suzor is an Australian Research Council Future Fellow at Queensland University of Technology School of Law and QUT’s Digital Media Research Centre. He is also a Chief Investigator of the ARC Centre of Excellence for Automated Decision-Making and Society and a member of the Oversight Board.
E-mail: n [dot] suzor [at] qut [dot] edu [dot] au

Acknowledgments

The authors acknowledge the support of the Australian Research Council Centre of Excellence for Automated Decision-Making and Society (CE200100005) and an Australian Research Council Discovery Project grant (DP170100122).

Notes

4. In addition to moderator usernames, our keywords included derivatives of ‘ban’, ‘moderators’, ‘quarantine’, and ‘rules’. We found these keywords to be effective in returning explicit discussion about each subreddit’s rules and their approach to enforcement.
References


Incels on Reddit: A study in social norms and decentralised moderation


Incels on Reddit: A study in social norms and decentralised moderation


r/AgainstHateSubreddits, 2020. “r/AgainstHateSubreddits — Open letter to Steve Huffman and the Board of Directors of Reddit, Inc — If you believe in standing up to hate and supporting black lives, you need to act,” Reddit, at https://www.reddit.com/r/AgainstHateSubreddits/comments/gyvqem/open_letter_to_steve_huffman_and_the_board_of/.
Incels on Reddit: A study in social norms and decentralised moderation


r/IncelsWithoutHate, 2017. “This is what Chad Gets while we sit in our rooms Coping,” at https://rareddit.com/r/IncelsWithoutHate/comments/e3g5jy/, accessed 11 May 2022.


Yishan Wong, 2014. “Every man is responsible for his own soul,” *r/blog*, at
Incels on Reddit: A study in social norms and decentralised moderation

https://www.reddit.com/r/blog/comments/2foivo/every_man_is_responsible_for_his_own_soul/, accessed 24 January 2022.

Editorial history

Received 29 March 2022; revised 9 May 2022; revised 10 May 2022; accepted 10 May 2022.

This paper is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Incels on Reddit: A study in social norms and decentralised moderation
by Rosalie Gillett and Nicolas Suzor.
First Monday, volume 27, number 6 (June 2022).
doi: https://dx.doi.org/10.5210/fm.v27i6.12575