“What are you doing, TikTok?” : How Marginalized Social Media Users Perceive, Theorize, and “Prove” Shadowbanning

DANIEL DELMONACO, University of Michigan, USA
SAMUEL MAYWORM, University of Michigan, USA
HIBBY THACH, University of Michigan, USA
JOSH GUBERMAN, University of Michigan, USA
AURELIA AUGUSTA, Carnegie Mellon University, USA
OLIVER L. HAIMSON, University of Michigan, USA

Shadowbanning is a unique content moderation strategy receiving recent media attention for the ways it impacts marginalized social media users and communities. Social media companies often deny this content moderation practice despite user experiences online. In this paper, we use qualitative surveys and interviews to understand how marginalized social media users make sense of shadowbanning, develop folk theories about shadowbanning, and attempt to prove its occurrence. We find that marginalized social media users collaboratively develop and test algorithmic folk theories to make sense of their unclear experiences with shadowbanning. Participants reported direct consequences of shadowbanning, including frustration, decreased engagement, the inability to post specific content, and potential financial implications. They reported holding negative perceptions of platforms where they experienced shadowbanning, sometimes attributing their shadowbans to platforms' deliberate suppression of marginalized users' content. Some marginalized social media users acted on their theories by adapting their social media behavior to avoid potential shadowbans. We contribute *collaborative algorithm investigation*: a new concept describing social media users' strategies of collaboratively developing and testing algorithmic folk theories. Finally, we present design and policy recommendations for addressing shadowbanning and its potential harms.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: content moderation, social media, marginalization, shadowbanning, algorithmic folks theories, collaborative algorithm investigation

ACM Reference Format:
Daniel Delmonaco, Samuel Mayworm, Hibby Thach, Josh Guberman, Aurelia Augusta, and Oliver L. Haimson. 2024. “What are you doing, TikTok?” : How Marginalized Social Media Users Perceive, Theorize, and “Prove” Shadowbanning. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 154 (April 2024), 39 pages. https://doi.org/10.1145/3637431

Authors' addresses: Daniel Delmonaco, dan.delmonaco@rutgers.edu, University of Michigan, Ann Arbor, Michigan, USA; Samuel Mayworm, mayworms@umich.edu, University of Michigan, Ann Arbor, Michigan, USA; Hibby Thach, hibby@umich.edu, University of Michigan, Ann Arbor, Michigan, USA; Josh Guberman, guberman@umich.edu, University of Michigan, Ann Arbor, Michigan, USA; Aurelia Augusta, aurelia@aeva.dev, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Oliver L. Haimson, haimson@umich.edu, University of Michigan, Ann Arbor, Michigan, USA.

2573-0142/2024/4-ART154 $15.00
https://doi.org/10.1145/3637431

1 INTRODUCTION

> *"Twitter release me from twitter shadowban!!! I won't talk about suckin and fuckin nomore. I promise that was 2020 behavior!"* [22]

Musical artist Cardi B tweeted this plea to Twitter in January 2021 when it appeared some of her content was not shown in followers' Twitter feeds. While Cardi B is a celebrity, she has a lot in common with many social media users, particularly marginalized social media users, who notice drops in engagement with content and believe shadowbanning is the culprit. In this case, Cardi B theorized that the shadowbanning occurred because she had posted about "suckin and fuckin." The platform seemed to deprioritize her more adult content from appearing in her followers' feeds. Twitter denies that shadowbanning occurs on the platform [52, 137], yet marginalized Twitter users sometimes believe they experience this specific type of content moderation. In this paper we consider participant experiences with and beliefs surrounding shadowbanning to understand how this surreptitious form of content moderation impacts platform users, especially those with marginalized identities.

We address the following research questions in this paper:

> RQ1. How do marginalized social media users who experienced shadowbanning make sense of shadowbanning?
> RQ2. What are the impacts of suspected shadowbanning on marginalized social media users who believe they experienced shadowbanning?

To answer these research questions, we conducted qualitative surveys (n = 71) and semi-structured interviews (n = 24) with marginalized social media users who experienced suspected or confirmed content moderation in the last year. We situate shadowbanning within existing algorithmic folk theory literature [38–40, 43, 82] because participants often developed their theories individually and collectively (within online communities) to explain discrepancies in engagement numbers and other signs of suspected shadowbanning. To understand how marginalized users make sense of perceived experiences of shadowbanning within the context of an unseen algorithm [104] and demurring social media companies [23, 114, 131], we explore and discuss the resourcefulness with which marginalized users produce and co-produce algorithmic folk theories related to shadowbanning. Developing shadowbanning-related algorithmic folks theories also represented participants' resourcefulness within a given situation in which they attempted to overcome the information gaps and asymmetries inherent to working with a blackboxed algorithm. We make the following contributions in this paper:

1. Provide an empirical understanding of marginalized social media users' experiences with, perceptions about, and consequences participants reported in relation to shadowbanning
2. Introduce the concept of *collaborative algorithm investigation* in which social media users collectively investigate and test algorithmic folk theories.
3. Present design and policy recommendations to increase transparency related to shadowbanning and communication with social media users, particularly marginalized social media users, who suspect they faced shadowbanning.

Further, in Appendix A, we detail how social media platforms have framed and discussed shadowbanning based on public statements.

2 BACKGROUND AND RELATED WORK

2.1 An overview of shadowbanning in popular and academic literature

> *"Where did the concept of 'shadow banning' come from?"*
> *"What is shadow banning and why does it deserve our attention?"*

"What is a shadowban and why does it matter?"

These provocative popular press article titles point to a recent journalistic goal to understand and define shadowbanning [2, 28, 110]. In this section we synthesize past shadowbanning discussions in popular press and academic literature, to ground participants' experiences in the surrounding context. In our results, we will discuss participants' confusion over shadowbanning and difficulties defining this specific type of content moderation.

Despite its more recent popularity in media coverage and online, shadowbanning is not a new phenomenon. Cole [28] outlined an extensive and thoroughly-researched history of shadowbanning and traced its origins to early Bulletin Board System (BBS) servers in the 1980s where administrators flagged certain users and restricted their access to platform features [142]. Shadowbanning gained increased notoriety and popularity as a term when the 45th President of the United States tweeted about alleged censorship of conservative tweets via shadowbanning [124]. This claim became a conservative talking point in the United States [94] and Twitter was quick to deny this practice [26, 52]. The only confirmation of censorship claims seemed to come from Jack Dorsey, former CEO of Twitter, in a 2018 testimony before the U.S. Congress in which he admitted Twitter was "unfairly" filtering certain accounts in auto-complete search results and "latest results" within Twitter's search feature [26]. After acquiring Twitter in 2022, Elon Musk drew further attention to "visibility filtering" during the release of the "Twitter Files" [84]; though Musk announced that Twitter would develop a "true account status" tool informing users if they are shadowbanned, this feature has not been released [100]. Conservative claims of suppression on social media are not supported by research [13, 46, 61, 95, 102, 124] and while conservatives do have more content removed than others, this tends to be because they post more content that violates site policies [68]. Shadowbanning and other acts of content moderation actually disproportionately impact marginalized social media users [68, 122].

Major platforms criticized for shadowbanning, including Instagram, Twitter, TikTok, and Facebook, released various statements distancing themselves from shadowbanning and/or claiming it did not exist on their respective platforms. In Table 4 (Appendix A), we present some of the platform responses to shadowbanning or similar instances of content deprioritization. Platforms rarely use the word shadowbanning in their statements, but TikTok, Twitter, YouTube, and Instagram have outright denied the practice by name [52, 88, 98, 114, 132]. When major shadowbanning events occur, such as TikTok's suppression of #BlackLivesMatter content, platforms tend to frame these issues as errors or other technical mistakes [26, 29, 75, 85, 106]. Social media algorithms are inherently opaque, and as users and researchers we do not know how exactly they filter and remove or prioritize content [44, 58, 59, 108, 139]. This opacity allows platforms to wave off any possible issue as a "bug" in the algorithm (i.e., as unintentional). Le Merrer et al. [86] analyzed shadowbanning instances and found that perceived shadowbans on Twitter were unlikely to be system-wide bugs as the platform claimed. Blanket statements about platform bugs neglect users' individual shadowbanning experiences and the resultant harms they face.

One content moderation strategy that some platforms (namely Facebook, Instagram, YouTube, and TikTok) do admit to is deprioritizing specific content [74, 98, 135]. Platforms do not use the term "shadowban" to describe this type of moderation [74, 98, 135]. In an explanation of their content moderation strategy, Instagram admitted that certain content would not appear on the Explore page and certain hashtag pages if it was deemed inappropriate for the "broader community" [74]. Mark Zuckerberg, Meta CEO, explained part of Facebook's content moderation strategy, saying, "We train AI systems to detect borderline content so we can distribute that content less" [145]. Facebook patented an automated content moderation process system that would "weed through user content" [77]. YouTube also admits to "demoting" videos containing "borderline" content from their content recommendation system, stating that the demotion system prevents “borderline” or “low-quality content” from being recommended to viewers while still being allowed on the platform [62]; YouTube’s implementation of the “de-recommendation” system has also resulted in flagged videos experiencing decreased share counts on the platform [20]. In a public statement addressing certain content’s absence from user feeds, TikTok released a news brief recognizing the possibility that the site may create a “filter bubble” where a user might see homogeneous content on their “For You Page” due to the recommendation system’s algorithm [135]. Platforms create filter bubbles or algorithmically deprioritize content deemed “inappropriate” or “borderline” presumably with the intention of protecting users from viewing possibly harmful content, but users seem to experience these practices as shadowbanning [8].

Shadowbanning and de-prioritizing content are examples of what Gillespie [60] calls “reduction as a form of content moderation” in which, rather than removing content outright, a platform instead demotes it, often using algorithmic means to determine which content to reduce. Reduction is a quiet strategy for platforms, and is substantially less risky for them politically than removing content [60]. Yet people do not trust platforms to do reduction work fairly and thoughtfully, primarily because users do not have sufficient transparency into who makes these decisions and how [60]. Content reduction highlights platforms’ tremendous power to choose what the public sees, which often happens in a way that is inequitable for marginalized communities [60]. Recently, many marginalized users perceived shadowbanning to include added layers of misogyny, racism, and other discrimination due to how users’ identities related to removed content [17, 31, 53]. In this paper we discuss how participants often tied experiences of shadowbanning to their marginalized identities.

Shadowbanning’s impact might especially target marginalized people because content related to their identities falls into content moderation borderline areas, such as content that may be considered sexual or nudity [10, 96]. Recent studies have found that marginalized groups (e.g., gender and racial minorities) perceive that they are disproportionately targeted for shadowbanning [42, 103, 109]. Cotter [34] suggested that shadowbanning is one example of how platforms deploy “black box gaslighting” in which platforms use their authority over their algorithms to undermine user experiences and observations about platform algorithms. Are [4, 6, 8–10] has written extensively on the topic of shadowbanning, among other forms of content moderation, in academic spaces and via her blog “Blogger On Pole”. Are [6] experienced shadowbanning on Instagram when she discovered that her pole dancing content was absent from the “Explore” page on others’ Instagram accounts. Drawing from personal experiences with shadowbanning, Are [8] described the “Shadowban Cycle.” In this process, Instagram and other platforms:

- Fail to remove harmful content such as hate speech and harassment
- Face public pressure to address these issues
- Target content such as pole dancing that platforms allege violates their content policies
- Prove that they went for an “easy” target of moderation rather than addressing the actual harassment of hate speech complained about in the first place [8].

The Shadowban Cycle presents shadowbanning as a wrong but easy solution to public criticism [8]. “Borderline” content, such as pole dancing, is algorithmically suppressed [9, 10, 71] in the name of protecting users, but appears to harm marginalized users instead of stopping hate speech and harassment.

As it continues to become a more popular term for describing often nontraditional content moderation experiences, academic literature has begun to address shadowbanning and social media user perceptions thereof. Myers West [101] provided an early definition of shadowbanning in 2018 as a phenomenon in which “content is made invisible to other users without actually being removed entirely." Participants in Myers West's study theorized shadowbanning as a type of content moderation where content was not actively removed by platforms [101]. In a 2022 report, Nicholas stated that shadowbanning means "to limit or eliminate the exposure of a user, or content or material posted by a user, to other users of the social media Internet site through any means, regardless of whether the action is determined by an individual or an algorithm, and regardless of whether the action is readily apparent to a user" [103]. Several studies have empirically examined social media shadowbanning and users' perceptions of it. In one study, some Facebook users felt stifled and silenced by the platform's shadowbanning since this content moderation strategy blocked users from appearing in others' Facebook News Feeds [73]. In a study on content moderation and transparency, Suzor et al. [127] reported that some surveyed users suspected that shadowbanning explained certain experiences of removed or hidden content without any notice from the platform. While a Twitter shadowban audit study found that shadowbans occurred rarely, political tweets and tweets with offensive content were more likely to be shown to smaller audiences [76].

While limited academic discussions about shadowbanning have been published in Media Studies and Communication venues [6, 8, 9, 34, 42, 60, 76, 101, 127], shadowbanning has been relatively absent from human-computer interaction (HCI) and social computing literature. In this paper, we situate shadowbanning in existing HCI research on content moderation and algorithmic folk theories, examine marginalized social media users' experiences with shadowbanning, and consider implications for HCI.

2.2 Content Moderation and Marginalized Groups

Previous social media research finds that social media users with marginalized identities face unique harms and challenges [68] due to major social media platforms' content moderation practices [16, 42, 116]. Content moderation on platforms occurs both algorithmically and by human content moderators, both of which can lead to detrimental impacts for marginalized social media users and their online experiences [56, 90]. Users' disagreements with platforms' moderation decisions, including decisions that disproportionately impact marginalized users, highlight the conflict between user experiences on platforms and content moderation policies in the form of contested platform governance [128]. For example, Tumblr's Not Safe for Work (NSFW) ban in 2018 was contested and criticized by Tumblr users for disadvantaging many marginalized users on the platform, particularly LGBTQ+ users and sex workers [67, 128], and, when implemented, prompted a massive decrease in users and migration away from the site. Both experiencing and anticipating content removals can have negative ramifications for online content posters from marginalized backgrounds, along with stakeholders reliant on content [12].

As previously discussed, platforms face criticism for shadowbanning and other unfair, often invisible [134] content moderation practices based on race [53, 64, 103], sexuality [51, 72, 109, 114], gender [6, 31, 57, 69, 109], and disability [17, 85, 109, 111, 140]. Certain social media platforms have been found to disproportionately shadowban marginalized users and their content. For example, Facebook and Instagram have been found to disproportionately shadowban women who post content (including hashtags) relating to fitness, pole dancing, and sex work [5, 30, 31, 112] on their platforms. TikTok has also been shown to restrict the visibility of content posted by disabled creators and LGBTQ+ creators from their For You Page, along with videos including rainbow flag emoji or hashtags related to fatness or disability [17, 51, 72, 85, 111, 114]. Black TikTok creators have also reported decreased follower and view counts after posting content related to race, experiences with racism, and the Black Lives Matter (BLM) movement [53]. For marginalized individuals, targeted shadowbanning leads some to feel isolated from online communities and important information, and some content (e.g., related to sex and disability) does not reach the intended audience who might benefit from it [103, 140]. Users relying on social media for their income also face direct financial consequences when shadowbanned, as they lose potential customers when their content does not appear on newsfeeds [32]. Sex workers report frequent shadowbanning instances despite adhering to community guidelines, which directly reduces their reach to potential clients and subscribers [15, 36, 48]. When posts about one's own identity are shadowbanned, platforms send a clear message about who has a place on the platform.

Alternatives to traditional content moderation present possibilities for more inclusive platforms that do not disproportionately harm marginalized people. For example, consent-based approaches to sexual content rather than blanket removals of potentially nude or sexually explicit content might prevent platforms from erasing sexuality and miscategorizing certain content as lewd [123]. Online communities face unique content moderation challenges, and community-based moderation strategies can include both methods that punish certain behaviors and encourage those found desirable [118]. Tailored moderation approaches, like including healthcare professionals in moderating mental health online communities, may better support certain marginalized users than blanket moderation practices [115]. Increased transparency of moderation practices is another popular recommendation for improving users' content moderation experiences [68, 79, 103, 127]. For example, transparency in the form of additional information about post removals can reduce future post removals by notifying users why a certain post faced removal [79]. Educating users about post removals also might improve user attitudes about fairness of removals and increase likelihood to post again [78]. Users might develop their own strategies for improving their platform experiences, such as creating Twitter blocklists to address online harassment, when a platform's content moderation systems fail to adequately support users [80].

Content facing removal often falls into content moderation "gray areas" which both algorithmic and moderator methods cannot easily categorize as "right" or "wrong" [68]. Haimson et al. [68] argued that moderation practices should embrace these gray areas in their moderation practices rather than forcing content to fit into strict permissible or removable categories on platforms. Marginalized users might also feel compelled to behave or act a certain way, even if harmful, due to the types of bodies and behaviors content moderation practices emphasize as "normal" [47].

2.3 Algorithmic Content Moderation

Algorithmic content moderation can exacerbate harms inflicted by platform content moderation processes by obscuring moderation practices, decreasing fairness perceptions, and further politicizing moderation decisions [63]. Due to shadowbanning's reported nature as a purposefully opaque form of content moderation [16, 18], algorithmic content moderation deserves particular attention in our attempt to understand social media users' shadowbanning experiences. Repeated instances of shadowbanning against marginalized groups such as LGBTQ+ and Black users point to potential algorithmic biases [18] built into these recommendation systems that operate the TikTok For You Page, Instagram Explore page, and Twitter or Facebook newsfeeds. Concepts like algorithmic misogynoir [90] and platformed racism [91] point to how discrimination and harm against marginalized people are codified in large-scale social media platforms' content moderation practices. Despite the previous harms mentioned, algorithmic content moderation methods might prove useful in specific online communities, such as pro-eating disorder groups, in the flagging and removal of potentially harmful and triggering content [24, 25].

Platforms justify increased automated content moderation practices with claims that there is no other way to keep up with their enormity and to moderate at scale [59]. Algorithmic content moderation allows platforms to consolidate specific content and then take an active role in patrolling communications about specific topics [27]. Algorithmic moderation fundamentally changes content moderation to a rule-based system rather than a "series of discrete decisions" [144], and content moderation requires platforms to make trade-offs between goals like cooperation and abuse prevention [81]. Suzor et al. [127] argued for increased content moderation transparency at a “systems level” for platforms to demystify content moderation practices at scale. Increased machine learning techniques in content moderation might also be most effective in supporting human moderators rather than replacing them entirely [59]. Even content moderation appeal systems, however, might not meet social media users’ needs and do not increase fairness, accountability, user control, or trust in the content moderation process [139]. Algorithmic content moderation is specifically relevant to the present study because our findings and previous shadowbanning claims [8, 18] tie shadowbanning directly to algorithmic content moderation.

2.4 Algorithmic Folk Theories

When interacting with technological systems, users might seek to make sense of phenomena by developing personal folk theories [54, 83]. Toff and Nielsen [136] defined folk theories as “the culturally available symbolic resources that people use to make sense of their own media and information practices.” We situate perceptions about shadowbanning in existing folk theory literature to understand how marginalized social media users develop theories about this mysterious type of observed and experienced content moderation. Bucher [19] suggested an “algorithmic imaginary” where users meet algorithms and form theories about algorithms and their functions in order to understand algorithms’ social power. Folk theories are one possible lens for thinking about this imaginary and the impact of users’ algorithmic understandings on their behavior. Previous folk theory literature in HCI situates users’ understandings of content moderation [59] and how social media feeds function within the framework of folk theory [39, 40, 43, 82]. Folk theories in HCI about algorithmically-driven systems, such as social media platform content moderation, center platform users’ experiences to understand how they develop algorithmic awareness and how this impacts their behavior [40]. Making users aware of algorithmic social media processes, such as the Facebook News Feed, led to their creation of folk theories similar to those already aware of social media algorithms [43]. Increased algorithmic awareness can increase engagement on platforms and create feelings of control in users [44]. Users develop algorithmic folk theories from many different information sources and these theories inform their self-presentation online [38].

Social media algorithms rely on user behavior to curate platform feeds, but often there can be mismatch between user goals and the algorithm’s functioning [108]. YouTube creators found the platform’s tiered governance system too ambiguous and improperly communicated to them, so they developed theories about why their content faced demonetization [21]. TikTok users formed folk theories about how the platform’s algorithm suppressed certain social identities such as race and ethnicity, class, LGBTQ+ identities, body size, physical appearance, disability, and affiliation with certain political and social justice groups [82]. This “identity strainer theory” [82] indicates a level of algorithmic awareness in users who created folk theories around their individual identities and the platform’s algorithm. Karizat et al. [82] and Simpson & Semaan found algorithmic folk theories motivated behaviors to “coach” or “domesticate” the algorithm to prioritize or deprioritize certain content on the platform’s For You Page; however, Simpson & Semaan also found that TikTok users cannot fully “domesticate” the algorithm into always prioritizing agreeable content, resulting in conflicts between the platform and its users’ “personal moral economies” [120]. The algorithmic clustering of identity-related content can also result in the erasure of queer content that does not cleanly fit into individual identity-based categories; Simpson & Semaan found that TikTok’s algorithm can reduce the visibility of “non-normative” queer content [121], while DeVito found that TikTok’s algorithmic content clustering can reduce visibility for multiply-marginalized trans users’ content (such as videos posted by trans women of color about their intersecting marginalized experiences) [37]. In another troubling context, Moran et al. [97] found that Instagram users create folk theories to evade content moderation when spreading anti-vaccination misinformation.

Increased understanding and theorization of the algorithms behind social media platforms can lead users to resist the harms perpetrated by platforms in a number of ways. One type of protest, algorithmic resistance, is users' resistance to the algorithms driving social media systems within the bounds of the platform [45, 141]. Velkova and Kaun [141] suggested algorithmic resistance as a corrective to social media platforms' algorithmic power and the detrimental impacts such algorithms can have on users. Resistance to algorithmic systems can provide a sense of agency and upset algorithmic systems' dominance [45]. For example, in the face of pole dancers facing shadowbanning on Instagram, one user found switching her profile's gender to male remedied her previously declining engagement numbers [30]. Other examples include queer TikTok users resisting algorithmic silencing by reposting queer content previously removed from the platform [121], or transfeminine TikTok users using folk theories about the platform to navigate and resist the suppression of videos related to transfemininity [37]. Alternative social media platforms might also serve as a type of resistance against shadowbanning and the detrimental practices some major platforms employ [7, 66, 143]. TikTok users who believed the platform's algorithm suppressed certain content related to their identities participated in both individual and collective acts of algorithmic resistance [82]. Rumored algorithmic changes to Twitter led users to develop folk theories about the platform and attempt to resist these changes [39]. Some of these acts of resistance highlight platforms' major shortcomings. In pro-eating disorder communities, users avoided certain hashtags and found other ways to avoid algorithmic content moderation in order to continue posting and circulating potentially harmful content [55]. On TikTok, some young users evade content moderation through shared strategies by creating a collective "algorithmic folklore" around discriminatory content moderation practices [1, 87].

In the case of resistance and shadowbanning, some users develop their own algorithmic folk theories for working against the algorithm to avoid or reverse shadowbanning. For example, Bain [11] presented tips to get "un-shadowbanned" on TikTok, including clearing your TikTok cache and deleting then redownloading the app. As another example, Simpson and Semaan described TikTok users collaboratively amplifying the visibility of content by LGBTQIA+ and BIPOC creators that they suspected were suppressed on the platform. Others attribute shadowbanning-like consequences to hashtag usage or behaving "like a bot" [92, 130].

Since shadowbanning remains purposefully opaque and ambiguous, users also desire methods for determining if they are in fact shadowbanned. Shadowbanning "tests" and other tools claim to determine if shadowbanning has in fact occurred when users experience less engagement or other assumed shadowbanning consequences [50]. Guides Don't Delete Art, an artist advocacy group, released a guide for artists facing shadowbanning of their art which includes methods such as "self censorship" via pixelation or cover-up and removing certain hashtags from their posts [41, 70]. Many of these attempts to detect shadowbanning fall under "everyday algorithmic auditing" in which users attempt to detect and understand potentially problematic interactions with algorithmic systems [119]. In this paper, we discuss users' folk theories of shadowbanning and their attempts to understand, prove, and possibly avoid shadowbanning in relation to platform algorithms. We focus on the construction and contents of user folk theories, as opposed to the veracity of said theories. Next, we describe the methods that we used to answer our research questions.

3 METHODS

Our results about shadowbanning come from analyzing two data sources: 1) Qualitative surveys with 71 users who had recently experienced content moderation and 2) 24 semi-structured interviews with users who had recently experienced content moderation. In this section we describe our data collection and analysis methods for both surveys and interviews. All aspects of this study were reviewed and deemed exempt from oversight by our university's Institutional Review Board (IRB) [1].

3.1 Surveys

3.1.1 Data Collection. Out of 326 total survey respondents, 71 of them (21.78%) who previously experienced content moderation in some form also reported also experiencing shadowbanning. We focus on this subset of 71 survey participants in this paper. The larger survey from which we took this subsection of responses about shadowbanning asked participants about their content moderation experiences with an oversampling for participants with marginalized identities. Participants were primarily recruited using panel survey companies Qualtrics (n = 37) and Prolific (n = 31), but we gathered some responses via social media sites such as Twitter, Facebook, and Reddit by posting in online communities relevant to marginalized populations and through our extended personal social media networks (n = 3). Participants were eligible for the survey if they were over the age of 18, lived in the United States, and had experienced content or account removals in the past year. Both the Prolific and Qualtrics surveys oversampled for racial and ethnic minorities, LGBTQ+ people, trans and/or nonbinary people, and participants across the political spectrum. The survey included 35 questions: 10 demographic questions (demographics reported in Table 1)), 14 open-ended questions, and 11 multiple choice questions about content moderation experiences. Table 2 indicates the platforms on which participants claimed to experience shadowbanning. Participants could choose multiple gender, sexuality, and race/ethnicity options, so percentages add up to greater than 100%. The survey was pilot tested and workshopped with colleagues, after which we made changes to survey structure and questions. We piloted a sample using Prolific (n = 20) and carefully read through responses to gauge if participants correctly interpreted questions. We then fully deployed the survey using Prolific and Qualtrics. While collecting responses, we carefully monitored survey responses to ensure data quality. We removed all responses where participants did not complete the survey or if text appeared to be gibberish or computer-generated. Qualtrics compensated participants directly for completion of the survey in accordance with Qualtrics' compensation policy (in either cash or cash-equivalent "points"); we paid Qualtrics $8 per completed survey response. Prolific participants received compensation at a rate at or above $12 per hour. Participants recruited via social media were entered into a drawing for a $50 gift card.

3.1.2 Data Analysis. The data on shadowbanning came from the 71 participants who answered "Yes" when asked "Within the last year, have you personally experienced shadowbanning on a social media site?" We compiled and analyzed these 71 participants' responses to all survey questions related to shadowbanning experiences. The first author read through all shadowbanning responses and conducted open coding [33]. During this process, the first author and fourth author developed a codebook and used axial coding [33] to group codes into larger themes. The first author discussed codes with the other authors. After discussion we revised the codebook and themes before coding all data. Themes in our codebook and discussed in this paper include: definitions of shadowbanning, finding proof of shadowbanning, consequences of suspected shadowbanning, perceptions about platforms, and shadowbanning and marginalized identities. The first author and the fourth author completed a second round of coding on the full dataset noting whether the code did or did not occur

[1] At our institution, interview and survey studies are generally deemed exempt from IRB oversight; IRB oversight is usually reserved for medical trials and studies with more in-depth or long-term interactions with participants. However, we took substantial precautions to practice ethical research and ensure that we protected participants' data, such as giving all participants anonymized participant numbers for audio recording and reporting purposes, restricting data access to the research team and transcribers bound to a confidentiality agreement, storing data on the research team's secure password-protected computers and secure servers, and deleting all interview audio recordings once transcripts were created and verified.

Table 1. Survey Participant Demographics

	# of Participants (total n = 71)	**% of Participants** (total n = 71)
Age		
18-24	26	36.62%
25-34	25	35.21%
35-44	14	19.72%
45-54	4	5.63%
55-64	2	2.82%
Gender		
Man	30	42.25%
Woman	36	50.70%
Nonbinary	8	11.27%
Sexuality		
Straight	47	66.20%
Bisexual	10	14.08%
Gay	6	8.45%
Lesbian	4	5.63%
Queer	4	5.63%
Pansexual	2	2.82%
Asexual	1	1.41%
Race/Ethnicity		
White	33	46.48%
Black or African American	22	30.99%
Hispanic or Latino	14	19.72%
Asian	7	9.86%
American Indian or Alaska Native	4	5.63%
Middle Eastern	2	2.82%

Participants could choose multiple gender, sexuality, and race/ethnicity options, so percentages add up to greater than 100%.

for each participant's response. The first author,the fourth author, and the last author discussed and resolved all issues of disagreement. The final dataset was annotated with 0/1 indicators denoting whether each code applied to each participant's data.

3.2 Interviews

3.2.1 Data Collection. The first, second, and third authors conducted semi-structured interviews with 24 participants. 23 interviews were conducted remotely over Zoom and recorded for audio transcription, and 1 interview with a deaf participant was conducted through text over email. We recruited participants in three ways: 1. Contacting participants from the previous survey who expressed interest in a follow-up interview (n = 6); 2. Using our personal social media accounts on Twitter to promote the study and share a screening survey (n = 6); 3. Using a research recruiting service (User Interviews) and its internal screening survey (n = 12). The screening survey asked participants if they experienced content or account removals on social media in the past year, and we oversampled for people from marginalized communities (racial, ethnic, sexual, and gender minorities). The screening survey asked participants about their most memorable content moderation experience and their age, gender, race/ethnicity, and LGBTQ+ status. Interview participant demographics are reported in Table 3.

Table 2. Reported Shadowbanning on Platforms by Survey Participants

Platform	**# of Reported Shadow-banning Instances** (total n = 71)
Facebook	28
Instagram	19
Twitter	15
TikTok	11
Reddit	4
YouTube	4
Tumblr	3
WhatsApp	3
Discord	2
Snapchat	2
Pinterest	2
LinkedIn	1
Telegram	1
Amino	1
MeetMe	1

Table 3. Interview Participant Demographics

ID	**Age**	**Gender**	**LGBTQ+**	**Race/Ethnicity**
P1	28	Woman	Yes	Black
P2	31	Nonbinary, agender	Yes	Middle Eastern
P3	27	Nonbinary, gender neutral	Yes	Asian
P4	24	Man	Yes	White
P5	40	Woman	Yes	White
P6	24	Nonbinary	Yes	Mixed
P7	21	Nonbinary	Yes	Asian
P8	26	Nonbinary	Yes	Asian
P9	24	Nonbinary	Yes	Asian
P10	28	Man	Yes	Asian
P11	23	Woman	Yes	Asian
P12	28	Woman	No	Hispanic/Latinx
P13	31	Woman	No	Hispanic/Latinx
P14	18	Nonbinary	Yes	Hispanic/Latinx
P15	31	Woman	No	Black
P16	44	Woman	No	Hispanic/Latinx
P17	36	Woman	No	Black
P18	40	Man	No	Hispanic/Latinx
P19	33	Woman	Yes	Black
P20	21	Woman	No	Native Hawaiian/Pacific Islander
P21	20	Woman	Yes	Asian
P22	31	Man	Did not disclose	Black
P23	23	Man	No	Asian
P24	22	Man	Yes	Asian

We completed a total of 24 interviews. We conducted 23 using Zoom and one using email to accommodate the participant's accessibility needs. All interviews were recorded and transcribed. Interviews lasted an average of 52 minutes (*sd* = 11 minutes; range: 38-84 minutes). After completing the informed consent process, we asked participants about their content or account removals and how their content moderation experience impacted them. We also asked about perceptions of content moderation policies and possible improvements or alternatives for current content moderation practices and community guidelines. Most of the results in this paper come from answers to the specific question "Have you heard of shadowbanning?" and follow up questions depending on the answer (e.g., "How did you know that shadowbanning was happening?"). Only two interview participants (8.33%) had never heard of shadowbanning. The remaining participants had varying experiences with and opinions about shadowbanning. Participants received $30 for participating in this interview study.

3.2.2 Data Analysis. The first three authors conducted open coding [33] using Atlas.ti. The first three authors began by all coding the same transcript to develop a codebook through open coding [33]. Our entire research team then met to discuss codes and refine the codebook. Once we reached agreement on all codes and themes and their meanings, we then coded interviews separately and discussed any disagreements throughout the process. The first author then used directed coding [33] on all interview data related to shadowbanning. The first author discussed this directed coding with the other authors throughout the process and incorporated their feedback into coding and development of themes around shadowbanning. Similar themes to those in the survey data emerged about shadowbanning: definitions of shadowbanning, finding proof of shadowbanning, consequences of suspected shadowbanning, perceptions about platforms, and shadowbanning and marginalized identities. Again, we are interested in users' experiences and subsequent sense-making activities related to these themes, rather than, for example, whether user definitions or suspicions of shadowbanning matched any objectifiable realities corresponding to said themes [126].

3.3 Positionality

This paper's authors collectively represent a broad spectrum of marginalized identities and lived experiences. The team includes multiple authors with lived experience across queer, trans, and nonbinary identities, multiple authors who are racial minorities and/or represent a mixed-race background, and multiple disabled authors. The authors are all marginalized social media users who are familiar with (and have experienced) the various kinds of identity-based harm marginalized users on the internet often face. The authors' broad range of marginalized identities benefits their collective ability to build rapport with participants during interviews and to interpret and understand participants' experiences with marginalization. Each of the authors also holds privilege in some ways, especially as highly-educated researchers at US universities. We took a reflexive approach to acknowledging our privilege and understanding how facets of our identities differ from participants and may sometimes limit our interpretation of their experiences.

4 RESULTS

In our results, we first describe participants'[2] confusion about shadowbanning - both its meaning and if it occurred to them. We then present the theories participants formed about shadowbanning, including how to "prove" it occurred via followers and decreased engagement. Third, we detail

[2] Interview participants are denoted by I# and survey participants are denoted by P#. Survey participant numbers refer to our full survey dataset (n = 326) which is why participant numbers are higher than the subset of 71 (who had experienced shadowbanning) focused on in this paper

consequences of shadowbanning and perceptions about platforms' roles in shadowbanning. Lastly, we describe how shadowbanning disproportionately harms marginalized social media users.

4.1 What is shadowbanning?

4.1.1 Confusion about shadowbanning's meaning. Our survey and interview data indicate that shadowbanning does not have a clear definition, which creates difficulty when social media users try to determine if shadowbanning occurred. We present results indicating participants' uncertainty around the term and differing shadowbanning definitions.

Platforms do not alert users that content became shadowbanned. Without a clear notification mechanism, it is difficult to know when or if it occurred. For instance, P191, a white nonbinary person, said, "*I never got a notice or notification. The content would simply be removed and I wouldn't be allowed to post on my account; sometimes for a couple hours, sometimes days at a time.*" Without a notification, P191 determined for themselves that they were shadowbanned. Rather than shadowbanning, this experience appears to be an account suspension or blockage. We cannot definitively say that P191 did not face shadowbanning, but it is important to acknowledge that they believe it occurred.

Lack of communication about shadowbanning was a source of confusion for some participants. When asked about shadowbanning, I2, a Middle Eastern nonbinary person, said, "*People will be saying, 'I think I've been shadow banned? Can anybody see this tweet?' And some other times people might be, 'I've been put in Twitter jail,' or 'I can't tweet for like 24 hours.' So I do wonder, is this a version of Twitter jail, but they just don't tell you?*" I2's suggestion that shadowbanning might be a type of "Twitter jail" could expand possible definitions of shadowbanning to include not just content suppression but also invisible temporary account bans. Lack of communication and transparency from platforms further obscures shadowbanning's definition and how participants think about this type of content moderation.

Shadowbanning and online communities arose as another area of confusion within defining and determining shadowbanning. P167, a Black man, responded, "*I guess on Reddit, based on my history on my old account, I couldn't participate in other subs with my opinions. In Facebook, I just simply disagreed with a moderator's opinion and they shadowbanned me there.*" In both of these instances, P167 attributed shadowbanning with limiting participation in an online community at the hands of a moderator. Unlike most other experiences of shadowbanning in this study, specific content was not suppressed but instead a social media user was excluded from participation in a specific group. P65, a Black woman, similarly said, "*Group members could not comment as the group moderator adjusted the settings to allow only the group moderator to make posts only.*" If these experiences are considered shadowbanning, this indicates that shadowbanning may include instances when online community moderators suppress content, not just when a platform itself suppresses content.

4.1.2 Forming theories about shadowbanning. Without a clear definition or understanding of shadowbanning, participants formed folk theories about these unique and unclear suppressions of content they experienced. Participants developed their algorithmic folk theories to explain how opaque and seemingly mysterious social media algorithms function to curate their feeds and shape their online experiences. In the case of shadowbanning, folk theories provided a way for participants to make sense of a specific type of content moderation which, according to participants, often suppresses marginalized groups' content. Participants attempted to make sense of a situation they viewed on social media by categorizing it as shadowbanning, and thus by contributing to a definition of shadowbanning.

Participants' most frequent folk theorization of shadowbanning was **decreased engagement with social media content that was disproportionate to previous engagement with their posted content**. For example, I9, an Asian nonbinary person, said,

> *There'll be times when I feel like, "Oh, people are seeing a lot of my posts..." Just based on likes or comments, I know a ton of people are seeing something. But other times, I would feel like, "Oh, I've posted a bunch of things in the last few days," like, these things have either maybe one like or no one has interacted with it. And so I always just kind of assumed maybe it was timing or maybe people, my friends, were less active on Facebook at that time or they just weren't as interested in that content. But actually, I'm not sure. Maybe that's kind of shadowbanning happening where my posts are really intentionally not being shown to people.*

While discussing experiences with lack of engagement with their content, I9 arrived at shadowbanning as a possible explanation for this lack of engagement with content. I9, however, was not certain they experienced shadowbanning or that shadowbanning is the right term for this incident. There are possible other explanations that participants theorized, **such as timing of the post or type of content**. I23, a South Asian (Indian) man, proposed the following definition and theories about shadowbanning:

> *Basically, your posts are not being visible to other people. Like you don't know that you're banned. But I mean, basically, other people can't see your posts and stories. Because it has happened to a few friends of mine. I guess. I've heard from a friend. I don't think it happened to me. I mean, I wouldn't know if I was shadowbanned, because that's the whole point of it.*

I23 did not experience shadowbanning personally, at least that he knew of, but formed this definition from experiences shared by friends. I23 also contributed the theory that **one purpose of shadowbanning is to limit the spread of certain content without officially blocking or removing the content on social media. Based on this logic, users will not realize they were shadowbanned if the platform's shadowbanning efforts are successful**. This further complicates attempts to define shadowbanning and users' abilities to determine if it occurred, which leads people to form folk theories about shadowbanning.

Participants who formed theories about what shadowbanning is also sometimes took this theorization a step further to theorize ways to prove that shadowbanning occurred. For example, I24, a South Asian (Indian) man, proposed:

> *I don't think there's always even a way to get unshadowbanned. I know people have brought up that you could think of a subreddit or a bot you could interact with and it will tell you your shadowbans. And I guess they check that by seeing a test account or a dummy account. And if they can see your posts being presented in the summary or some overview If you can see it then they can tell you're not shadowbanned. But if they don't see anything when you're posting something then they would tell you you're shadowbanned.*

According to I24, social media users developed bots or subreddits to perform an algorithmic audit of specific posts that would determine whether one's account had been shadowbanned. Since platforms claim that shadowbanning does not occur, there is not a traditional appeal process or avenue for reinstating moderated content. Participants were most concerned with determining whether or not their account and/or specific content faced shadowbanning.

But how do people attempt to confirm whether shadowbanning is actually happening? For participants in our study, **relying on feedback from social media followers and monitoring engagement with accounts or content were the most common methods among participants**

for determining shadowbanning. In the following section we discuss these two strategies in detail.

4.2 Have I been shadowbanned?

Participants theorized two main methods for determining shadowbanning: 1. Friends and followers confirming shadowbanning; 2. Monitoring changes in engagement. Confirmation from others often manifested as followers confirming whether or not they could see a posted video in their feeds. To monitor engagement, participants would often compare engagement numbers of similar posts across a period of time. Drastic drops in engagement from one video to the next would often lead participants to blame shadowbanning.

4.2.1 Friends/followers confirm shadowbanning. Confirming with followers or friends on social media was a frequent strategy for participants to determine if shadowbanning occurred. This usually meant participants suspected they were shadowbanned, and then recruited followers for assistance or were alerted of the shadowbanning by their followers. Without clear definitions from platforms or definitive mechanisms for determining shadowbanning, users are left to theorize for themselves if shadowbanning occurred to their accounts or content. For participants in our study, social media followers and friends acted as co-investigators with participants. P39, a Black woman, said, "*I realized I was shadowbanned after some mutuals mentioned not seeing my post.*" Since there was no notification from the platform, P39 could not confirm for herself that the post was shadowbanned.

Other participants discussed content creators and social media users specifically asking friends or followers to confirm whether they could see specific content. I23, a South Asian (Indian) man, discussed friends' experiences with shadowbanning:

> *So they got to know that they were shadowbanned, because usually people engage with them. And they put such stories or posts, and people react to their story... But they didn't have that coming up. So they contacted their other friends. They're like, "Hey, what's going on?" And they'll [the friends] be like, "Oh, yeah, we can't see any of your stories and posts."*

I23's friends realized, based on reduced engagement levels, that some type of moderation or suppression occurred with their social justice-related content. This participant's social justice content involved criticisms of the Hindutva Indian government's treatment of Muslims and LGBTQIA+ people. When asked about shadowbanned content, P244, a white man, described, "*I know they shadowbanned it because no one reacted to it, and no way no one would react to something like this. I tagged friends to make sure they could see it.*" P244 tagged specific friends to ensure that they saw the specific post. P244 suspected shadowbanning, although it is possible the tagged individuals chose not to interact with his post. I23 and P244's responses highlight how relying on others to confirm shadowbanning is one strategy people use when faced with substantial uncertainty about this specific type of content moderation. Reliance on others to prove shadowbanning, in lieu of notification from the platform, shifts the burden to users. Users rely on the kindness of friends and often strangers to confirm, but not everyone has the capacity or reach to use this strategy effectively.

Determining shadowbanning via confirmation by followers relies on other social media users' reciprocity. For example, I23 described an instance when mutuals reached out about a lack of engagement on her post, which alerted her to potential shadowbanning. Without followers' interventions, I23 might not have realized anything happened to her content. When discussing shadowbanning experiences, I21, an Asian American woman, described being on the other side of this reciprocal support:

> *They'll be people I follow, and I'll see them posting things like, "Oh, I think I'm shadowbanned, like, can you see this video? Can you let me know?" And then those will just like, come up randomly and I'll be like, "Oh, I saw your videos. So you should be fine."*

Here, I21 acted as an engaged and benevolent follower and confirmed to those who suspected shadowbanning that she could in fact see their content in her feed. Unlike other types of content moderation where users receive a notification from platforms and possible avenues for appeal, there are no definitive notices that shadowbanning occurred. Thus, social media users must work collaboratively to investigate suspected algorithmic content suppression.

4.2.2 Monitoring engagement with content. In addition to the reliance on followers for shadowbanning confirmation, some participants determined shadowbanning through social media engagement statistics. This often meant comparing content posted at different times and recognizing an abnormal lack of engagement when compared to previous engagement levels. The type of engagement varied across platforms but often led participants to naming shadowbanning as the reason for irregularities in video views or like counts. TikTok was the most mentioned platform where users observed or deployed this strategy.

Often, participants considered a lack of views or likes on certain content to be definitive proof that shadowbanning did in fact occur. P143, a biracial woman, responded, "*My videos on TikTok would be up for hours with zero views, and I have a good amount of followers and am not private. My video was simply not being shown.*" Based on P143's past TikTok posts and knowledge of the platform's algorithm behind the For You Page, she found shadowbanning to be the only explanation for her video's reduced engagement. When making this assessment, P143 considered amount of time since posting, amount of followers, and account privacy measures. I7, a Mixed (Chinese and Hispanic/Latinx) nonbinary person, similarly arrived at shadowbanning as an explanation for lower views when posting about their trans identity and trans health-related content. I7 said:

> *A couple of times I was answering people's questions, but since TikTok didn't like my content, they didn't push it to the For You Page like they have my other content. And so content that I was making with resources that people had specifically asked for, had gotten shadowbanned, and had gotten, like, 12 views on them. And I averaged around, like, 107 views, like, 180 views per video? So like, 11 is like, "Oh, cool, okay. What are you doing, TikTok?"*

I7 considered specific view counts to determine whether their content had been shadowbanned. Both P143 and I7 noted a disparity between the number of views different videos received, and cited these disparities as evidence of shadowbanning. They also blamed the specific platform (TikTok) for this lack of engagement, which we discuss further later in this paper.

Participants prevalently mentioned TikTok in their shadowbanning experiences, specifically a lack of views and the importance of the For You Page. The For You Page serves as the main landing page for TikTok in which videos are algorithmically curated for users based on previous activity on the platform. Several participants identified absence from the For You Page as the specific type of shadowbanning they experienced, and attributed the abnormal reduction in video views to this absence. P166, a white woman, answered, "*My videos are no longer being shown on the For You Page, and are barely reaching any of my followers; I've noticed a HUGE drop in views on my videos and people have even commented that they haven't seen me in a while.*" Platform specifics, such as the TikTok For You Page, can create an online environment where shadowbanning is both more noticeable and more consequential. Creators on TikTok rely on the platform's algorithms to appear on the For You Page of other users. Users of the platform also may feel less control over their digital selves [120] due to the platform's affordances. If there is a noticeable decrease in engagement, participants are likely to blame it on the feature that determines popularity of content, the For You Page and its algorithm in this instance.

Participants discovered shadowbanning via lack of engagement on other platforms, although not as often as on TikTok. For example, P168, a white man, said:

> *I noticed a significant reduction in the amount of likes and retweets (and general interactions) I was receiving particularly on Twitter only to find out that I was shadowbanned. I did not show up in searches or anything for a couple of weeks.*

On Twitter, engagement took the form of likes and retweets for P168. P168 also described not appearing in searches on Twitter, which no other participant mentioned as a component of shadowbanning. It is unclear how exactly appearing in searches fits into shadowbanning, which emphasizes the confusion around this social media phenomenon. Likes and retweets were the main unit of measure in P168's example while other participants who discussed engagement on TikTok were more concerned with views. The engagement parameters participants relied on for determining shadowbanning were context dependent. Participants used similar strategies but different measures across platforms.

4.3 Consequences of suspected shadowbanning

Participants in both our survey and interviews reported direct consequences of shadowbanning. These consequences included frustration, loss of followers, decreased platform usage, and potential financial implications. Consequences predominantly arose from decreased engagement with content and/or an inability to post specific content due to shadowbanning.

Several participants expressed frustration, sadness, and other negative feelings as a result of shadowbanning. P189, a white man, expressed, "*I felt frustrated because I did not expect something like that to happen to my account whatsoever.*" In certain instances, participants responded with decreased account activity or deleting their account altogether. P245, an Asian man, said, "*It was a very bad experience as I was so annoyed that I thought to delete my account.*" When noticing shadowbanning or other irregularities online that one might label shadowbanning, a response like P245's to disengage with the platform might seem like the user's only option. With no definitive "proof" and no mechanism for contacting or appealing the platform, users feel as if nothing can be done about shadowbanning. I20, a Native Hawaiian/Pacific Islander woman, responded:

> *Recently I've just been spending a bit more time on TikTok than I'd like, but I've been inspired to create more creative content, but the thing is because my account is now shadowbanned I don't think I could ever really post on that account specifically... There just wouldn't be much fulfillment. Obviously, I can post for myself, but it's nice to share my content with the world, and knowing that that account is not gonna happen is kinda sad*

I20 felt discouraged from using TikTok due to suspected shadowbanning of her content. The lack of engagement caused by suspected shadowbanning removed some of the fulfillment she felt when posting content on the platform and the lack of this external validation via views, likes, comments, and shares on TikTok resulted in negativity.

Some of these negative consequences caused participants to reevaluate their place in a specific online community or on a social media platform. I20 felt deprived of a space she came to enjoy as a creative outlet. I6, a mixed race (Chinese and Hispanic/Latinx) nonbinary person, had a similar relationship and experience with TikTok, although they came to an almost opposite conclusion with these negative feelings. I6 said:

> *"Made me mad" is an understatement. But it made me feel a little bit discouraged that, like, my body is still being "dinged" by an algorithm that is arbitrarily defined. And it*

> *makes me go "Oh, should I just not post trans content anymore? Should I just like stop posting on TikTok?" And I was like, "Well, no!" Because people are clearly, like, "We need to see more of this." And so I think it more lit a fire under my ass to be, like, "I will keep doing this then, and I will keep fighting it." And, like, it would suck if my account went away? But if that were to happen, I probably would just start passively consuming media again.*

In this instance, I6 posted content related to their trans identity and experienced anger when some of this content appeared to be shadowbanned. I6 evaluated their place on TikTok and considered stopping production of trans-related content. Instead of deleting their account or stopping, they decided to keep posting despite the shadowbanning. Although some might be similarly empowered to keep posting content they believe a platform is attempting to suppress, not everyone can afford to have the same response.

Another consequence of shadowbanning that participants identified was potential financial implications. For example, P117, a Hispanic/Latino woman, said "*My engagement had dropped abruptly and no one could see me, stunting my business.*" She faced a material consequence due to a decrease in engagement. When one's income relies on social media, shadowbanning can directly impact one's reach and earning potential.

Sex workers are one group who became increasingly reliant on social media platforms for income and engagement, especially in the midst of the COVID-19 pandemic and the 2017 passage of the Allow States and Victims to Fight Online Sex Trafficking Act and Stop Enabling Sex Traffickers Act (SESTA-FOSTA) in the United States [16, 122]. SESTA-FOSTA caused many internet platforms to remove sex workers' previously supported online content due to fears of liability for hosting content under the law [16]. I12, a Hispanic/Latinx woman, responded:

> *Yeah, I definitely have seen posts from sex workers that mentioned, you know, being shadowbanned affects them just because they're getting less engagement. They're getting less people going to their websites, and purchasing their services, their videos. So while I don't talk to sex workers, personally, the ones that I do follow on Twitter, have mentioned that it has affected them greatly because of that.*

I12 presents sex workers as a particular group harmed by shadowbanning. I12 continued, "*I know that with Facebook and Twitter and all that where just them talking about 'I have an OnlyFans' will get you shadowbanned.*" Sex workers face disproportionate amounts of shadowbanning and are impacted more intensely and more frequently that other social media users [16]. Discussions of sex work allowable within community guidelines, such as mentioning OnlyFans, could directly lead to shadowbanning in order to keep content off platforms. Sex workers who also identify as activists, organizers, or protesters additionally report financial losses, disruption of movements, inability to access mutual aid efforts, and restriction from social media marketing tools that non-sexworkers use on platforms [16]. There is a direct monetary consequence when sex workers' content is suppressed and potential subscribers or clients do not see their content. I12's response also raises concerns shared by other participants that specific types of content or accounts are subject to shadowbanning. We discuss this further in the next section.

4.4 Perceptions about platforms and their role in shadowbanning

Some participants blamed shadowbanning on social media platforms and their possible ulterior motives for suppressing certain types of content. As previously discussed, platforms deny shadowbanning, yet social media users' experiences indicate that something unusual is happening to their content. The lack of information about what participants name as shadowbanning leads them to blame social media companies and theorize possible explanations for shadowbanning. When asked why she was shadowbanned, P89, a biracial woman, responded, "*I am not sure, but there is evidence to support that TikTok does this to punish creators and maintain an agenda.*" TikTok has previously denied shadowbanning allegations [114] but did admit to suppression of certain content from marginalized users [17, 51]. P89 presented shadowbanning as a result of TikTok's alleged agenda and as a punitive measure against certain creators. Several participants attributed shadowbanning to a bias held by the platform. In reference to shadowbanning, P238, a white woman, said, "*I believe that the social media platforms have an agenda, particularly Twitter, and they don't even attempt to hide it. They have political bias.*" According to P238, the bias is specifically political and content is shadowbanned when is does not fit Twitter's alleged agenda.

Not all participants expressed certainty in the supposed platform agendas or the reasoning for shadowbanning specific content and people. I15, a Black woman, responded, "*I know that companies will never admit this, but like, what is the truth? Like, are you shadowbanning queer people? Or... plus sized people or anything like that? They're not going to admit it, but I would like to know.*" I15 recognized the reality that most likely platforms will not officially admit to shadowbanning. I15 also presented the possibility that platforms' goals of shadowbanning are to suppress certain people or bodies without officially removing their content or removing them from the platform.

The nature of shadowbanning leaves much room for confusion and speculation by social media users. Denying shadowbanning while users experience and claim to prove shadowbanning leads to folk theory creation in order to explain this obscure, indirect, and observable type of content moderation. All of this murkiness related to shadowbanning led some participants to consider why platforms participate in this practice. I9, an Asian (Chinese American) nonbinary person, said:

> *Yeah, there's also just other stuff that I always just vaguely attributed to the algorithm, whatever that means. Just thinking, "Oh, I used to see so many posts from this person in my feed and I haven't seen it for a while." And then I would go to their profile and then I'll see, "Oh, they have been posting actively all this time. I'm just not seeing their posts." So I was never really sure what that was. I thought maybe, "Oh Facebook algorithm just is being fickle and they think I'm more interested in these other people's posts." But yeah, it is possible maybe it's something to do with what things those people post that Facebook doesn't want me to see those posts anymore.*

I9 mentions that shadowbanning might occur algorithmically but introduces enough uncertainty that it is possible a platform like Facebook might prioritize certain content while deemphasizing and shadowbanning other content. I9's response also suggests that with shadowbanning as a form of content moderation, platforms arbitrate which content users might see. When platforms take on this role surreptitiously using undisclosed curatorial algorithms for social media feeds, participants appeared to assume ill intent and expressed their concerns that shadowbanning perpetuated certain platform agendas.

4.5 Shadowbanning and Marginalized Identities

Participants attributed certain instances of shadowbanning to the suppression of content created by people with marginalized identities, such as queer people and non-white people. According to some participants, discussing content related to marginalized identities, either their own or those of others, led to shadowbanning. For example, P99, a mixed race (Middle Eastern and white) nonbinary person attributed their shadowbanning to "*being a Jew with a vagina and a voice.*" P89, a mixed race (American Indian/Alaska Native and Hispanic/Latino) nonbinary person, reported, "*I was shadowbanned on TikTok following a video I made in support of the Black Lives Matter movement.*" P99 and P89 experienced shadowbanning and sought an explanation, which led to their hypotheses that the shadowbanning was related to religion, gender, and support of certain social justice movements.

Participants across different marginalized and intersecting identities reported experiencing shadowbanning. When asked about shadowbanning, I12, a Hispanic/Latinx woman, summarized, "*I think anybody that isn't a white cishet person, honestly, you know, Black, Indigenous people, people of color, posts are often taken off, I think, more than other groups. I know sex workers posts are often taken away, and they're often shadowbanned as well, not just on Twitter and Instagram. So yeah, basically, anyone that isn't white and cishet, or I would consider to be a marginalized group on social media.*" According to I12, anyone who is not white, cisgender, and heterosexual can face shadowbanning and its detrimental effects. As in previous work on marginalized people and content moderation experiences [68], participants with certain marginalized identities found their content related to these identities and/or social justice topics were removed by platforms without seeming to violate community guidelines.

Other participants spoke strongly about feeling silenced by social media platforms due to shadowbanning. I3, an Asian (Korean American) nonbinary person, said:

> *I definitely have heard of shadowbanning. And I do have strong opinions about it because it's been weaponized against many of my queer peers. People in my age group on the social media platforms that I use, having their voices restricted and muted because they're queer or marginalized, right? They're just being actively oppressed by shadowbanning. And that's just a fact. I think that the scary thing about shadowbanning is that it becomes very impossible to put the word out and you're essentially being stripped of your rights to use social media as it was intended.*

I3 claimed that shadowbanning was an "active" form of oppression against queer people on social media that limited this group's rights when using social media. The nature of shadowbanning presents a new understanding of social media "rights," since users are not officially banned or unable to post certain content. In the case of shadowbanning, participants are still able to use social media platforms, but their potential reach and engagement is limited.

A few participants discussed potential strategies users, especially those with marginalized identities, might use to avoid shadowbanning. For example, I7, a South Asian (Indian) nonbinary person, shared:

> *I think that even I, when I first got community guideline violation, made a video being like, "I was trying to show off my top surgery scars, and TikTok took it down. What the heck!" And that was the first time that trans creators were like, "I've tried to make videos about this, but those videos get shadowbanned. Here are some of my tips." And I was like, "Oh, cool." … People even will call TikTok "the clock app" as to not trigger an algorithm. There are certain ways you type captions, certain ways you type your description of videos, certain ways you talk in videos, to make sure that TikTok doesn't auto-flag you. So people were like, "I got flagged on TikTok for a community guideline violation for this reason. I don't think I should have been, but here's what you can do if you're a creator like me." But then it would get shadowbanned because they had said "TikTok" and "community guideline violation."*

I7 experienced shadowbanning for sharing information about gender affirmation surgery and other trans-related content. This presents an interesting situation where an original video about top surgery was moderated, then the creator posted a response video about the potentially unfair moderation, and then the response video seemed to be algorithmically moderated via shadowbanning for calling out the platform. Content creators on TikTok and other platforms, especially those with marginalized identities, develop and share ways to avoid automated content moderation by avoiding certain words, hashtags, or other content components leading to shadowbanning or other types of moderation [3, 30]. I7 shared one such strategy - calling TikTok "the clock app." Another such example that I1, a Black woman, provided was the use of "yt people" instead of "white people" to avoid suspected shadowbanning when using the latter. Users also noticed that criticisms of TikTok were shadowbanned and developed new strategies for criticism that use different words that are not being algorithmically filtered.

Participants' experiences show how marginalized communities are adapting to avoid shadowbanning while discussing their identities, social justice, and other related topics – but should they have to? Currently, the onus falls on users to create strategies around content moderation. I15, a Black woman, suggested: "*I think first of all, all the stuff about shadowbanning and not promoting queer people, people of color, whatever, all of that just needs to stop period. That's definitely a large problem.*" In the following discussion, we present recommendations to alleviate this pressure put on users and ways to mitigate harms caused by shadowbanning.

5 DISCUSSION

In this Discussion, we situate participant folk theorizations about shadowbanning in existing literature and frame some of these theories and strategies as algorithmic resistance to platforms. We contribute *collaborative algorithm investigation* as a new framework for understanding collectively shared information about how platform algorithms function. We discuss outcomes of shadowbanning in relation to its impacts on participants. Finally, we make design and policy recommendations for addressing shadowbanning and the harm it causes to participants and other social media users.

5.1 Shadowbanning Folk Theorization, Algorithmic Resistance, and Collaborative Algorithm Investigation

Participants' increased awareness of shadowbanning led some to feel the need to prove that shadowbanning occurred to them. Shadowbanning purposefully remains "in the shadows" as a content moderation technique and suppresses users' content without notifying them. Platforms deny the existence of shadowbanning [52, 88, 98, 114], yet participants reported its existence and proposed methods for determining that shadowbanning occured. In what follows, we situate these shadowbanning experiences and possible methods for proving its occurrence within current folk theory literature. Shadowbanning folk theories have particular importance for marginalized people in our sample who described feeling that shadowbanning related to their marginalized identities. Participants discussed "the algorithm" and the algorithmic processes of platforms and their content moderation strategies with a level of certainty despite no official notification from platforms and without official knowledge of how the algorithm actually works. Their folk theorizations about shadowbanning often relied on information they learned from other users. For example, some participants knew about shadowbanning because content creators they followed talked about being shadowbanned. This type of collective theorization fits into Bishop's [14] theory of algorithmic gossip, in which knowledge about algorithms and their visibility is communally and socially informed. Algorithmic gossip is one way to spread "algorithmic lore" [14, 89] about how a platform's algorithms function without actual proprietary knowledge to confirm the algorithmic techniques used. In the unique case of shadowbanning, the spread of algorithmic gossip about shadowbanning led participants to direct action. For instance, in a sort of "algorithmic audit," [119], participants used what they gathered via algorithmic gossip about shadowbanning to identify if it occurred or try to stop it with techniques they learned about elsewhere online, such as monitoring view counts or asking followers to confirm if content appeared in the followers' feeds. Collective gossip and knowledge sharing led to collective actions.

Marginalized social media users employ collective action related to shadowbanning as a form of algorithmic resistance against platforms [45, 141]. Developing ways to evade detection by refraining from certain hashtags resists the hidden nature of shadowbanning. The algorithmic gossip [14] that highlights shadowbanning and brings light to this content moderation type might also function as a resistance measure against platforms' power and black box gaslighting [34]. We introduce the concept of *collaborative algorithm investigation* to describe social media users' willingness to investigate and test each others' algorithmic folk theories and report findings to one another. The participants quoted in section 4.2.1 each describe examples where they provided or received assistance to/from other marginalized social media users, as a way of collectively attempting to understand how algorithmic shadowbanning may be impacting them. Collaborative algorithm investigation draws from Bishop's concept of algorithmic gossip, which describes the collaborative process in which social media users "formulate and sustain algorithmic expertise" [14]. Collaborative algorithm investigation extends algorithmic gossip to highlight the collective, charitable, reciprocal nature participants expressed when attempting to identify, prove, and/or evade shadowbanning and other forms of content suppression. Reliance on others for determining and possibly exposing or preventing shadowbanning arose out of necessity and a desire to make sense of irregularities in social media engagement.

Participants' algorithmic lore about shadowbanning often included collaborative algorithm investigation. Through exploring not only the content of shadowbanning folk theories users create, but also *how* they create these theories, we found users employing various forms of collaboration and collective theory building to attempt to define and prove shadowbanning and it's occurrences. In the case of shadowbanning, collaborative algorithm investigation involved attempts to prove shadowbanning, such as commenting on others' posts if users could or could not view certain content. We consider collaborative algorithm investigation to be one type of resistance strategy against platforms. Users not only develop and spread folk theories but rely on one another to test and confirm or deny their theories. Users often strive to make sense of interactions with technologies despite a dearth of important contextual and causal information [65, 125, 126]. Similarly, we found that users collaboratively theorize their interactions with the platform, despite the platform remaining opaque and providing no additional information to aid in this sense-making process. In the absence of explanations from the platforms and within the context of opaque algorithms, users relied primarily on other users, and the spirit of reciprocity, to develop and test folk theories. Using folk theories and these strategies, however, led some participants to qualify responses about shadowbanning with hesitancy and uncertainty. Without clear statements from platforms, shadowbanning folk strategies and algorithmic gossip will always be tinged with uncertainty. The onus also unfairly shifts to users, who must learn about these folk theories and strategies to evade shadowbanning and deploy them without clear guidelines about how and why content is shadowbanned. Thus, collaborative algorithm investigation is an important concept for social computing researchers and social media platforms to understand, so that we can further study and theorize marginalized social media users' social media practices under uncertainty.

5.2 Outcomes of Shadowbanning

People with marginalized identities had some of the strongest reactions to suspected shadowbanning and expressed their feelings of being silenced for posting content related to these identities. Shadowbanning led users to feel discouraged and almost betrayed by the platform for not even validating this secretive content moderation they believe they experienced. Cotter's [34] use of shadowbanning as an example of black box gaslighting seems quite appropriate based on our findings. Shadowbanning denial by platforms (see Table 4, Appendix A) disregards social media users' lived experiences. However, as we have shown, users do notice problems with platform algorithms. Prior work (in non-academic outlets) has shown that shadowbanning and the alleged algorithmic processes behind it perpetuate racism [53], sexism [30, 31], ableism [17, 85, 111, 117], and discrimination against LGBTQ+ users [17, 51, 114]. Participants in this study confirmed these prior reports when they shared similar experiences about when their content related to their own marginalized identities or to social justice causes faced shadowbanning. Trans participants had content about trans healthcare or media featuring scars from top surgery shadowbanned. A participant criticizing anti-LGBTQIA+ policies in India felt purposefully silenced when this content was shadowbanned. Participants who posted about topics related to their marginalized identities and then faced shadowbanning seemed to receive a clear message from platforms: the platform purposefully created a space where those identities were not seen or heard from through shadowbanning. Whether or not platforms continue to deny shadowbanning, marginalized users believe it is happening and experience real impacts. These impacts included anger and frustration, discouragement, loss of financial opportunities, and loss of followers and other engagement forms. Some participants sought to find ways around shadowbanning but others channeled their discouragement and frustration differently by using platforms less or considering leaving altogether.

5.3 Design and Policy Implications

We present several recommendations for addressing shadowbanning and the harms that participants described. These recommendations fall into two major categories: design recommendations and policy recommendations. Our design recommendations focus on specific technical features and possible solutions that social media platforms might enact to combat the shadowbanning's detrimental effects experienced by participants. Policy recommendations instruct social media platforms how to update their current content moderation policies to improve users' online experiences related to possible shadowbanning. The authors developed these recommendations in a collaborative process with our research team based on categorizing and discussing survey and interview participants' experiences.

5.3.1 Design Recommendations: Increased Transparency and Communication. Participants lacked access to specific information from platforms about their metrics and created their theories about shadowbanning to make sense of the content moderation occurring. As noted above, platforms deny shadowbanning's existence [26, 32, 52, 98, 137], but users such as our participants believe they experience shadowbanning and similar types of content suppression on social media. Social media companies are not clear or transparent about the algorithms they use to determine which content to promote or suppress [74, 98, 135, 145]. We recommend platforms implement design changes to improve transparency so users do not need to rely on folk theories but can definitively prove that something odd is happening with their content engagement, and determine if platforms are shadowbanning. Access to definitive metrics about view counts and engagement over time for creators would allow them to better track potential dips in engagement due to shadowbanning. Some platforms offer some individual metrics, such as Tweet Analytics, or offer more advanced metrics tools to influencers and businesses. Third party apps also exist to monitor metrics. These strategies, however, require users to track or monitor metrics for individual posts or find third party apps to do this. Platforms building infrastructure for monitoring metrics across time alleviates some of the burden on users. Allowing users to track engagement metrics across similar content posted by others might also help to definitively prove shadowban or at least indicate that something suspicious is happening with the algorithm. For instance, two users with similar followings might be able to compare their videos in support of Black Lives Matter using #BLM to learn if one user's post has drastically lower engagement numbers. While platforms are unlikely to implement these strategies because they may surface unfavorable insights, transparency is nonetheless an important way forward, and would help to improve trust and goodwill between platforms and users.

By denying shadowbanning, platforms retroactively address the topic when instances of alleged shadowbanning gain public attention, instead of listening to user concerns about shadowbanning as it occurs [5, 26, 29]. Participants in our study discussed the lack of communication from platforms about shadowbanning. To address this, we propose design suggestions for platforms to increase opportunities for communication with platforms at the individual user level. Platforms could implement reporting mechanisms specifically for instances of suspected shadowbanning in addition to the usual post reporting features. Users suspect that shadowbanning occurs as a form of algorithmic content moderation [8, 18, 34]. Platforms might implement notification systems to tell users if a post was flagged for deprioritization in newsfeeds by algorithmic detection or due to reporting by other users. As part of increased communication, platforms can also implement reporting features for users to upload and explain why they believe shadowbanning is occurring, and even appeal the shadowbanning. These design changes would require platforms to interact directly with users rather than quietly suppressing content, and could help validate marginalized users' concerns at the individual level.

5.3.2 Policy Recommendations: Increased Transparency and Communication. Next, we recommend policy improvements for platforms to address suspected shadowbanning instances and better support the harm users face. Increased transparency of content moderation practices at both the individual and platform level could contribute to increased understanding and accountability of platform content moderation practices for social media users [127]. In the case of shadowbanning and black box gaslighting [34] by platforms, publicly acknowledging shadowbanning is the first policy recommendation we suggest. Platforms' hesitancy to name and validate shadowbanning neglects the very real experiences of users, particularly marginalized social media users like participants in our study. Admitting outright that shadowbanning occurs might not be a realistic expectation from platforms. However, validating user experiences and admitting that many, often marginalized, users face something akin to the practice of shadowbanning would seem to at least begin addressing the distrust and frustration users (such as the participants in this study) describe feeling when faced with shadowbanning.

Better communication with users about shadowbanning might also improve users' experiences on social media platforms. Many participants in our study reported that they would not know they are shadowbanned until being alerted by followers or developing their own theories and tests based on engagement. If platforms do believe that some content falls into moderation grey areas [68] and that this leads them to deprioritize users' content without fully removing it from the platform [6, 74, 93, 135], then increased communication would at least allow users to know their content fell into a grey area. Rather than secretly suppressing content, platforms should openly tell users if their content falls into a grey area or is deemed to be borderline content. Rather than forcing users to develop these theories of shadowbanning and possible resistance, platforms could communicate with users and definitively inform them why algorithmic content moderation tools flagged a certain post. For example, in response to shadowbanning pole dancing on Instagram [5, 7], the platform could have alerted users who posted pole dancing content that certain hashtags led to not be shown on the Explore page. This policy recommendation goes hand in hand with our design recommendations.

We also recommend that platforms create dedicated shadowbanning response teams or similar internal groups devoted to understanding this type of content moderation and its impacts on users. Devoting resources and time to developing shadowbanning response teams might force companies to acknowledge that their algorithms impact users in this specific way. A shadowbanning response team or similar entity at social media platforms could implement and address shadowbanning appeal processes, and also help with outreach to those harmed by shadowbanning. Amongst participants in our study, shadowbanning led to further distrust of platforms. We recommend that these shadowbanning teams reach out directly to those impacted previously by shadowbanning, or what they believe to be shadowbanning, with an emphasis on marginalized communities. Incorporating those who experienced shadowbanning into platform response efforts would validate these negative online experiences that platforms previously denied. In this effort platforms also must take shadowbanning folk theories as stemming from experiences of (often marginalized) users and potential biases of their algorithms rather than dismissing shadowbanning altogether.

5.4 Limitations

Shadowbanning as a type of content moderation is difficult to study because we cannot definitively prove its existence through qualitative research with social media users. In this paper, social media users' experiences led to their beliefs in shadowbanning. Participants thought they experienced shadowbanning and content suppression in this way. In some instances, they also believed they could prove its existence. We cannot definitively prove shadowbanning's existence without access to platform algorithms. As presented in Table 4 (Appendix A), platforms deny shadowbanning as a content moderation strategy, but this might be a claim they cannot fully confirm as platform algorithms impact user experiences in unforeseen ways [86]. The social media companies themselves might not actually know what their algorithms are doing to users' content [86]. There is also a gap between what actually occurs on social media and what social media users believe occurs [138]. Whether shadowbanning does or does not occur at the platform's technical level, participant perceptions about the platform and the algorithm are what matter to us as social media researchers [138]. In this paper, we only analyzed data from participants who answered "Yes" when we asked if they had personally experienced shadowbanning, excluding those who answered "I'm not sure." Given shadowbanning's ambiguous nature, engaging with social media users who are unsure if they have experienced it is an interesting area for future research.

5.5 Future Directions in Shadowbanning

The previous recommendations work "within" the structures of social media companies and would generally require social media companies to acknowledge shadowbanning or at least that users experience something akin to it, often related to their marginalized identities. In the immediate future while companies continue to deny shadowbanning [52, 88], resources such as Don't Delete Art's Resource Center [41]accept the current social media platforms as they are and work with users to address shadowbanning within this reality. The website hosts a gallery for artists to display art previously banned from social media platforms and discuss their content moderation experiences. There is also a Resource Center with tips about avoiding content moderation before posting artwork and an "Appeals" page with information on appealing content moderation decisions across different platforms. On a similar note, our research team is currently developing an online help center for those with marginalized identities to both demystify the content moderation process and provide strategies for appealing unfair or opaque content moderation, such as shadowbanning.

6 CONCLUSION

We examined marginalized people's experiences with shadowbanning on social media platforms, and described how they construct folk theories to make sense of these experiences. We contribute a revised definition for shadowbanning to better encompass participants' experiences and provide recommendations to demystify shadowbanning and help to address some of the inequities it involves. We extend previous folk theory literature in our discussion of participant experiences with shadowbanning and attempts to prove shadowbanning, and contribute the concept collaborative algorithm investigation to describe the collaborative testing of algorithmic folk theories about social media platforms. We then discuss the consequences of shadowbanning with particular focus on users with marginalized identities.

As platforms continue to deny shadowbanning, users will continue to distrust platforms because they deny people's online lived experiences. We hope this paper validates experiences of participants and other social media users facing shadowbanning and leads social media platforms to acknowledge shadowbanning and the harms it causes. Cardi B might not receive a clear answer or solution to her shadowbanning plea [22], but we imagine a future where platforms take concerns about shadowbanning seriously and address them directly.

ACKNOWLEDGMENTS

We thank our study participants for sharing their insights and experiences with us. We thank Tarleton Gillespie, Kristi Gamarel, Casey Pierce, and the members of the Community Research on Identity and Technology (CRIT) Lab at UMSI for their helpful feedback and comments on our work. We also thank our anonymous reviewers for their constructive comments that improved this work. This work was supported by the National Science Foundation grant #1942125.

REFERENCES

[1] Iretiolu Akinrinade and Joan Mukogosi. 2021. Strategic Knowledge. https://points.datasociety.net/strategic-knowledge-6bbddb3f0259

[2] Paula Akpan. 2020. What Is Shadow Banning & Why Does It Deserve Our Attention? https://www.bustle.com/life/what-is-shadow-banning-how-does-it-work

[3] Travis M. Andrews. 2020. Tinder, TikTok and more: Online activists are finding creative new ways to say Black Lives Matter. *Washington Post* (June 2020). https://www.washingtonpost.com/technology/2020/06/12/tiktok-tinder-twitter-bts-black-lives-matter/

[4] Carolina Are. 2019. Instagram Apologises To Pole Dancers About The Shadowban. https://bloggeronpole.com/2019/07/instagram-apologises-to-pole-dancers-about-the-shadowban/

[5] Caroline Are. 2019. Instagram Denies Censorship of Pole Dancers and Sex Workers. https://bloggeronpole.com/2019/07/instagram-denies-censorship-of-pole-dancers-and-sex-workers/

[6] Carolina Are. 2020. How Instagram's algorithm is censoring women and vulnerable users but helping online abusers. *Feminist Media Studies* 20, 5 (July 2020), 741–744. https://doi.org/10.1080/14680777.2020.1783805 Publisher: Routledge _eprint: https://doi.org/10.1080/14680777.2020.1783805.

[7] Carolina Are. 2021. Interview with social media platform Lips. http://bloggeronpole.com/2021/10/interview-with-social-media-platform-lips/

[8] Carolina Are. 2021. The Shadowban Cycle: an autoethnography of pole dancing, nudity and censorship on Instagram. *Feminist Media Studies* 0, 0 (May 2021), 1–18. https://doi.org/10.1080/14680777.2021.1928259 Publisher: Routledge _eprint: https://doi.org/10.1080/14680777.2021.1928259.

[9] Carolina Are. 2022. An autoethnography of automated powerlessness: lacking platform affordances in Instagram and TikTok account deletions. *Media, Culture & Society* (Dec. 2022), 01634437221140531. https://doi.org/10.1177/01634437221140531 Publisher: SAGE Publications Ltd.

[10] Carolina Are and Susanna Paasonen. 2021. Sex in the shadows of celebrity. *Porn Studies* 0, 0 (Sept. 2021), 1–9. https://doi.org/10.1080/23268743.2021.1974311 Publisher: Routledge _eprint: https://doi.org/10.1080/23268743.2021.1974311.

[11] Ellissa Bain. 2020. TikTok: How to get 'un-shadowbanned' – temporary ban can be reversed! https://www.hitc.com/en-gb/2020/07/01/tiktok-how-to-get-un-shadowbanned-temporary-ban-can-be-reversed/ Section: Trending.

[12] Anna Veronica Banchik. 2021. Disappearing acts: Content moderation and emergent practices to preserve at-risk human rights–related content. *New Media & Society* 23, 6 (June 2021), 1527–1544. https://doi.org/10.1177/1461444820912724 Publisher: SAGE Publications.

[13] Paul M. Barrett and J. Grant Sims. 2021. *False Accusation: The Unfounded Claim that Social Media Companies Censor Conservatives.* Technical Report. NYU Stern Center for Business and Human Rights. https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/6011e68dec2c7013d3caf3cb/1611785871154/NYU+False+Accusation+report_FINAL.pdf

[14] Sophie Bishop. 2019. Managing visibility on YouTube through algorithmic gossip. *New Media & Society* 21, 11-12 (Nov. 2019), 2589–2606. https://doi.org/10.1177/1461444819854731 Publisher: SAGE Publications.

[15] Danielle Blunt and Ariel Wolf. 2020. *Erased: The Impact of FOSTA-SESTA & the Removal of Backpage.* Technical Report. Hacking//Hustling.

[16] Danielle Blunt, Ariel Wolf, Emily Coombes, and Shanelle Mullin. 2020. Posting Into the Void: Studying the Impact of Shadowbanning on Sex Workers and Activists. https://hackinghustling.org/posting-into-the-void-content-

moderation/

[17] Elena Botella. 2019. TikTok Admits It Suppressed Videos by Disabled, Queer, and Fat Creators. *Slate* (Dec. 2019). https://slate.com/technology/2019/12/tiktok-disabled-users-videos-suppressed.html

[18] Annie Brown. 2021. Understanding The Technical And Societal Relationship Between Shadowbanning And Algorithmic Bias. https://www.forbes.com/sites/anniebrown/2021/10/27/understanding-the-technical-and-societal-relationship-between-shadowbanning-and-algorithmic-bias/ Section: AI.

[19] Taina Bucher. 2017. The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society* 20, 1 (Jan. 2017), 30–44. https://doi.org/10.1080/1369118X.2016.1154086

[20] Cody Buntain, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker. 2021. YouTube Recommendations and Effects on Sharing Across Online Social Platforms. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–26. https://doi.org/10.1145/3449085

[21] Robyn Caplan and Tarleton Gillespie. 2020. Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy. *Social Media + Society* 6, 2 (April 2020), 2056305120936636. https://doi.org/10.1177/2056305120936636 Publisher: SAGE Publications Ltd.

[22] Cardi B. 2021. Twitter release me from twitter shadowban!!! I won’t talk about suckin and fuckin nomore .I promise that was 2020 behavior! https://twitter.com/iamcardib/status/1345977698368200704

[23] Twitter Help Center. [n.d.]. Debunking Twitter myths. https://help.twitter.com/en/using-twitter/debunking-twitter-myths

[24] Stevie Chancellor, Yannis Kalantidis, Jessica A. Pater, Munmun De Choudhury, and David A. Shamma. 2017. Multimodal Classification of Moderated Online Pro-Eating Disorder Content. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI ’17)*. ACM, New York, NY, USA, 3213–3226. https://doi.org/10.1145/3025453.3025985 event-place: Denver, Colorado, USA.

[25] Stevie Chancellor, Zhiyuan (Jerry) Lin, and Munmun De Choudhury. 2016. "This Post Will Just Get Taken Down": Characterizing Removed Pro-Eating Disorder Social Media Content. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI ’16)*. ACM, New York, NY, USA, 1157–1162. https://doi.org/10.1145/2858036.2858248 event-place: San Jose, California, USA.

[26] CNBC Television. 2018. Twitter CEOJack Dorsey Testifies - Sept. 5, 2018. https://www.youtube.com/watch?v=41P9cbaWiBc

[27] Jennifer Cobbe. 2020. Algorithmic Censorship by Social Platforms: Power and Resistance. *Philosophy & Technology* (Oct. 2020). https://doi.org/10.1007/s13347-020-00429-0

[28] Samantha Cole. 2018. Where Did the Concept of ’Shadow Banning’ Come From? https://www.vice.com/en/article/a3q744/where-did-shadow-banning-come-from-trump-republicans-shadowbanned

[29] Charlotte Colombo. 2021. TikTok has apologized for a ’significant error’ after a video that suggested racial bias in its algorithm went viral. https://www.insider.com/tiktok-racism-algorithm-apology-creator-marketplace-ziggy-tyler-2021-7

[30] Jesselyn Cook. 2019. Instagram’s Shadow Ban On Vaguely ‘Inappropriate’ Content Is Plainly Sexist. https://www.huffpost.com/entry/instagram-shadow-ban-sexist_n_5cc72935e4b0537911491a4f Section: Women.

[31] Jesselyn Cook. 2019. Women Are Pretending To Be Men On Instagram To Avoid Sexist Censorship. https://www.huffpost.com/entry/women-are-pretending-to-be-men-on-instagram-to-avoid-sexist-censorship_n_5dd30f2be4b0263fbc99421e Section: Tech.

[32] Jesselyn Cook. 2020. Instagram’s CEO Says Shadow Banning ‘Is Not A Thing.’ That’s Not True. https://www.huffpost.com/entry/instagram-shadow-banning-is-real_n_5e555175c5b63b9c9ce434b0 Section: Politics.

[33] Juliet M. Corbin and Anselm L. Strauss. 2008. *Basics of qualitative research techniques and procedures for developing grounded theory* (3rd ed. ed.). Sage Publications, Inc., Los Angeles, Calif.

[34] Kelley Cotter. 2021. “Shadowbanning is not a thing”: black box gaslighting and the power to independently know and credibly critique algorithms. *Information, Communication & Society* 0, 0 (Oct. 2021), 1–18. https://doi.org/10.1080/1369118X.2021.1994624 Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2021.1994624.

[35] Antigone Davis and Amit Bhattacharyya. 2021. How Meta Addresses Bullying and Harassment. https://about.fb.com/news/2021/11/how-meta-addresses-bullying-harassment/

[36] Britt Dawson. 2020. Instagram’s problem with sex workers is nothing new. https://www.dazeddigital.com/science-tech/article/51515/1/instagram-problem-with-sex-workers-is-nothing-new-censorship Section: Science & Tech.

[37] Michael Ann DeVito. 2022. How Transfeminine TikTok Creators Navigate the Algorithmic Trap of Visibility Via Folk Theorization. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–31. https://doi.org/10.1145/3555105

[38] Michael Anne DeVito, Jeremy Birnholtz, Jeffery T. Hancock, Megan French, and Sunny Liu. 2018. How People Form Folk Theories of Social Media Feeds and What it Means for How We Study Self-Presentation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–12. https:

//doi.org/10.1145/3173574.3173694

[39] Michael Anne DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. "Algorithms ruin everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 3163–3174. https://doi.org/10.1145/3025453.3025659

[40] Michael Anne DeVito, Jeffrey T. Hancock, Megan French, Jeremy Birnholtz, Judd Antin, Karrie Karahalios, Stephanie Tong, and Irina Shklovski. 2018. The Algorithm and the User: How Can HCI Use Lay Understandings of Algorithmic Systems?. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–6. https://doi.org/10.1145/3170427.3186320

[41] Don't Delete Art. 2021. Resource Center. https://dontdelete.art/resource-center/

[42] Brooke Erin Duffy and Colten Meisner. 2022. Platform governance at the margins: Social media creators' experiences with algorithmic (in)visibility. *Media, Culture & Society* (July 2022), 01634437221111923. https://doi.org/10.1177/01634437221111923 Publisher: SAGE Publications Ltd.

[43] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I "like" it, then I hide it: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 2371–2382. https://doi.org/10.1145/2858036.2858494

[44] Motahhare Eslami, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. 2015. "I always assumed that I wasn't really that close to [her]": Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 153–162. https://doi.org/10.1145/2702123.2702556

[45] Nancy Ettlinger. 2018. Algorithmic affordances for productive resistance. *Big Data & Society* 5, 1 (Jan. 2018), 2053951718771399. https://doi.org/10.1177/2053951718771399 Publisher: SAGE Publications Ltd.

[46] Brian Feldman. 2018. Twitter Is Not 'Shadow Banning' Republicans. https://nymag.com/intelligencer/2018/07/twitter-is-not-shadow-banning-republicans.html

[47] Jessica L. Feuston, Alex S. Taylor, and Anne Marie Piper. 2020. Conformity of Eating Disorders through Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (May 2020), 40:1–40:28. https://doi.org/10.1145/3392845

[48] Juniper Fitzgerald and Jessie Sage. 2019. Shadowbans: Secret Policies Depriving Sex Workers of Income and Community. https://titsandsass.com/shadowbans-secret-policies-depriving-sex-workers-of-income-and-community/

[49] Instagram for Business. 2017. We understand users have experienced issues with our hastag search that c aused posts to not be surfaced. We are continuously working on improvements to our system with the resources available. https://www.facebook.com/instagramforbusiness/posts/1046447858817451

[50] Caroline Forsey. 2021. Instagram Shadowban Is Real: How to Test for & Prevent It. https://blog.hubspot.com/marketing/instagram-shadowban

[51] Chris Fox. 2020. TikTok admits restricting some LGBT hashtags. *BBC News* (Sept. 2020). https://www.bbc.com/news/technology-54102575

[52] Vijaya Gadde and Kayvon Beykpour. 2018. Setting the record straight on shadow banning. https://blog.twitter.com/en_us/topics/company/2018/Setting-the-record-straight-on-shadow-banning

[53] Meira Gebel. 2020. Black Creators Say TikTok Still Secretly Hides Their Content. https://www.digitaltrends.com/social-media/black-creators-claim-tiktok-still-secretly-blocking-content/

[54] Susan A. Gelman and Cristine H. Legare. 2011. Concepts and Folk Theories. *Annual Review of Anthropology* 40 (2011), 379–398. https://www.jstor.org/stable/41287739 Publisher: Annual Reviews.

[55] Ysabel Gerrard. 2018. Beyond the hashtag: Circumventing content moderation on social media. *New Media & Society* 20, 12 (Dec. 2018), 4492–4511. https://doi.org/10.1177/1461444818776611

[56] Ysabel Gerrard. 2020. Social media content moderation: six opportunities for feminist intervention. *Feminist Media Studies* 20, 5 (July 2020), 748–751. https://doi.org/10.1080/14680777.2020.1783807 Publisher: Routledge _eprint: https://doi.org/10.1080/14680777.2020.1783807.

[57] Ysabel Gerrard and Helen Thornham. 2020. Content moderation: Social media's sexist assemblages. *New Media & Society* 22, 7 (July 2020), 1266–1286. https://doi.org/10.1177/1461444820912540 Publisher: SAGE Publications.

[58] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media.* Yale University Press, New Haven.

[59] Tarleton Gillespie. 2020. Content moderation, AI, and the question of scale:. *Big Data & Society* (Aug. 2020). https://doi.org/10.1177/2053951720943234 Publisher: SAGE PublicationsSage UK: London, England.

[60] Tarleton Gillespie. 2022. Do Not Recommend? Reduction as a Form of Content Moderation. *Social Media + Society* 8, 3 (July 2022), 20563051221117552. https://doi.org/10.1177/20563051221117552 Publisher: SAGE Publications Ltd.

[61] Kayla Gogarty and Spencer Silva. 2020. A new study finds that Facebook is not censoring conservatives despite their repeated attacks. https://www.mediamatters.org/facebook/new-study-finds-facebook-not-censoring-conservatives-

despite-their-repeated-attacks

[62] Cristos Goodrow. 2021. On YouTube’s recommendation system. https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/

[63] Robert Gorwa, Reuben Binns, and Christian Katzenbach. 2020. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7, 1 (Jan. 2020), 2053951719897945. https://doi.org/10.1177/2053951719897945 Publisher: SAGE Publications Ltd.

[64] Julia Angwin Grassegger, Hannes. 2017. Facebook’s Secret Censorship Rules Protect White Men From Hate Speech But Not Black Children. https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms?token=dQhE1DjH5S-s5oiZGbyX7NoNfXm5VycL

[65] Keith Grint and Steve Woolgar. 1997. *The Machine at Work: Technology, Work and Organization* (1st edition ed.). Polity, Cambridge, UK ; Malden, MA : Blackwell Publishers.

[66] Oliver L. Haimson, Justin Buss, Zu Weinger, Denny L. Starks, Dykee Gorrell, and Briar Sweetbriar Baron. 2020. Trans Time: Safety, Privacy, and Content Warnings on a Transgender-Specific Social Media Site. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 124:1–124:27. https://doi.org/10.1145/3415195

[67] Oliver L. Haimson, Avery Dame-Griff, Elias Capello, and Zahari Richter. 2019. Tumblr was a trans technology: the meaning, importance, history, and future of trans technologies. *Feminist Media Studies* 21, 3 (2019), 345–361. https://doi.org/10.1080/14680777.2019.1678505 Publisher: Routledge _eprint: https://doi.org/10.1080/14680777.2019.1678505.

[68] Oliver L. Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 466:1–466:35. https://doi.org/10.1145/3479610

[69] Oliver L. Haimson, Tianxiao Liu, Ben Zefeng Zhang, and Shanley Corvite. 2021. The Online Authenticity Paradox: What Being "Authentic" on Social Media Means, and Barriers to Achieving It. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 423:1–423:18. https://doi.org/10.1145/3479567

[70] Gareth Harris. 2021. Censored? Shadowbanned? Deleted? Here is a guide for artists on social media. https://www.theartnewspaper.com/news/don-t-delete-art-social-media-censorship-guide

[71] Amelie Heldt. 2020. Borderline speech: caught in a free speech limbo? https://policyreview.info/articles/news/borderline-speech-caught-free-speech-limbo/1510

[72] Alex Hern. 2019. TikTok’s local moderation guidelines ban pro-LGBT content. *The Guardian* (Sept. 2019). https://www.theguardian.com/technology/2019/sep/26/tiktoks-local-moderation-guidelines-ban-pro-lgbt-content

[73] Monica Horten. 2021. *Algorithms Patrolling Content: Where’s the Harm?* SSRN Scholarly Paper ID 3792097. Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.3792097

[74] Instagram. 2021. Why are certain posts on Instagram not appearing in Explore and hashtag pages? | Instagram Help Center. https://help.instagram.com/613868662393739

[75] Instagram Comms. 2021. We know that some people are experiencing issues uploading and viewing stories. This is a widespread global technical issue not related to any particular topic and we’re fixing it right now. We’ll provide an update as soon as we can. https://twitter.com/InstagramComms/status/1390376354332487681

[76] Kokil Jaidka, Subhayan Mukerjee, and Yphtach Lelkes. 2023. Silenced on social media: the gatekeeping functions of shadowbans in the American Twitterverse. *Journal of Communication* (Jan. 2023), jqac050. https://doi.org/10.1093/joc/jqac050

[77] Cal Jeffrey. 2019. USPTO grants Facebook patent for automated shadow-banning system. https://www.techspot.com/news/80979-uspto-grants-facebook-patent-automated-shadow-banning-system.html

[78] Shagun Jhaver, Darren Scott Appling, Eric Gilbert, and Amy Bruckman. 2019. "Did You Suspect the Post Would be Removed?": Understanding User Reactions to Content Removals on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 192:1–192:33. https://doi.org/10.1145/3359294

[79] Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2019. Does Transparency in Moderation Really Matter?: User Behavior After Content Removal Explanations on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–27. https://doi.org/10.1145/3359252

[80] Shagun Jhaver, Sucheta Ghoshal, Amy Bruckman, and Eric Gilbert. 2018. Online Harassment and Content Moderation: The Case of Blocklists. *ACM Trans. Comput.-Hum. Interact.* 25, 2 (March 2018), 12:1–12:33. https://doi.org/10.1145/3185593

[81] Jialun Aaron Jiang, Peipei Nie, Jed R. Brubaker, and Casey Fiesler. 2022. A Trade-off-centered Framework of Content Moderation. (June 2022). https://doi.org/10.1145/3534929 arXiv:2206.03450 [cs].

[82] Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. 2021. Algorithmic Folk Theories and Identity: How TikTok Users Co-Produce Knowledge of Identity and Engage in Algorithmic Resistance. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 305:1–305:44. https://doi.org/10.1145/3476046

[83] Willett Kempton. 1986. Two Theories of Home Heat Control*. *Cognitive Science* 10, 1 (1986), 75–90. https://doi.org/10.1207/s15516709cog1001_3 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog1001_3.

[84] David Klepper, Barbara Ortutay, and Matt O'Brien. 2022. EXPLAINER: How Elon Musk is changing what you see on Twitter. https://apnews.com/article/elon-musk-twitter-inc-technology-europe-business-1b3d4266c5acdab47fc1c95fe8026590

[85] Chris Köver. 2019. Discrimination - TikTok curbed reach for people with disabilities. https://netzpolitik.org/2019/discrimination-tiktok-curbed-reach-for-people-with-disabilities/ Library Catalog: netzpolitik.org.

[86] Erwan Le Merrer, Benoît Morgan, and Gilles Trédan. 2021. Setting the Record Straighter on Shadow Banning. In *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*. 1–10. https://doi.org/10.1109/INFOCOM42981.2021.9488792 ISSN: 2641-9874.

[87] Amanda Lenhart and Kellie Owens. 2021. The Unseen Teen. https://datasociety.net/library/the-unseen-teen/ Publisher: Data & Society Research Institute.

[88] Jackie Lerm. 2020. I asked @mosseri this question, knowing full well how he was going to respond. There you have it guys. Again. Shadowbanning is not a thing. #SMSpouses https://t.co/LXGzGDjpZH. https://twitter.com/jackielerm/status/1231122961379340289

[89] Thomas WL MacDonald. 2021. "How it actually works": Algorithmic lore videos as market devices. *New Media & Society* (June 2021), 14614448211021404. https://doi.org/10.1177/14614448211021404 Publisher: SAGE Publications.

[90] Brandeis Marshall. 2021. *Algorithmic misogynoir in content moderation practice*. Technical Report. Heinrich-Böll-Stiftung. 17 pages.

[91] Ariadna Matamoros-Fernández. 2017. Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society* 20, 6 (June 2017), 930–946. https://doi.org/10.1080/1369118X.2017.1293130 Publisher: Routledge _eprint: https://doi.org/10.1080/1369118X.2017.1293130.

[92] Stacey McLachlan. 2021. Experiment: I Tried to Get Shadowbanned on Instagram. https://blog.hootsuite.com/experiment-i-tried-to-get-shadowbanned-on-instagram/

[93] Bryan Menegus. 2019. Facebook Patents Shadowbanning. https://gizmodo.com/facebook-patents-shadowbanning-1836411346

[94] Anna Merlan. 2020. How Shadowbanning Went from a Conspiracy Theory to a Selling Point. https://www.vice.com/en/article/v7gq4x/how-shadowbanning-went-from-a-conspiracy-theory-to-a-selling-point-v27n3

[95] Danaë Metaxa, Joon Sung Park, James A. Landay, and Jeff Hancock. 2019. Search Media and Elections: A Longitudinal Investigation of Political Search Results. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 129:1–129:17. https://doi.org/10.1145/3359231

[96] Callie Middlebrook. 2020. *The Grey Area: Instagram, Shadowbanning, and the Erasure of Marginalized Communities.* SSRN Scholarly Paper ID 3539721. Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.3539721

[97] Rachel E. Moran, Izzi Grasso, and Kolina Koltai. 2022. Folk Theories of Avoiding Content Moderation: How Vaccine-Opposed Influencers Amplify Vaccine Opposition on Instagram. *Social Media + Society* 8, 4 (Oct. 2022), 20563051221144252. https://doi.org/10.1177/20563051221144252 Publisher: SAGE Publications Ltd.

[98] Adam Mosseri. 2021. Shedding More Light on How Instagram Works. https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works

[99] Adam Mosseri. 2022. Account Status Update. https://www.instagram.com/reel/Cl34K-BAm3P/?utm_source=ig_embed&ig_rid=bdbd2f52-168b-4f5f-809d-d722cc18e1b1

[100] Elon Musk. 2022. Twitter Account Status/Shadowban Update. https://twitter.com/elonmusk/status/1601042125130371072?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1601042125130371072%7Ctwgr%5Ecd9c40c1d61926f0930c77752976acf1c1bf863c%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fhypebeast.com%2F2022%2F12%2Felon-musk-twitter-shadowban-account-update

[101] Sarah Myers West. 2018. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society* 20, 11 (Nov. 2018), 4366–4383. https://doi.org/10.1177/1461444818773059

[102] Casey Newton. 2019. The real bias on social networks isn't against conservatives. https://www.theverge.com/interface/2019/4/11/18305407/social-network-conservative-bias-twitter-facebook-ted-cruz

[103] Gabriel Nicholas. 2022. *Shedding Light on Shadowbanning*. Technical Report. Center for Democracy & Technology. 52 pages.

[104] Cathy O'Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (1st edition ed.). Crown, New York.

[105] onlinecensorship.org. [n.d.]. onlinecensorship.org — Submit Your Report. https://onlinecensorship.org/takedowns/new

[106] Vanessa Pappas and Kudzi Chikumbu. 2020. A message to our Black community. https://newsroom.tiktok.com/en-us/a-message-to-our-black-community?from=from_parent_docs

[107] Vanessa Pappas and Kudzi Chikumbu. 2020. Progress Report: How we're supporting Black communities and promoting diversity and inclusion. https://newsroom.tiktok.com/en-us/progress-report-how-were-supporting-black-communities-and-promoting-diversity-and-inclusion

[108] Emilee Rader and Rebecca Gray. 2015. Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 173–182. https://doi.org/10.1145/2702123.2702174

[109] Jess Rauchberg. 2022. #Shadowbanned: Queer, trans, and disabled creator responses to algorithmic oppression on TikTok. In *LGBTQ Digital Cultures: A Global Perspective*. Routledge. Google-Books-ID: fYFZEAAAQBAJ.

[110] Andreas Rekdal. 2021. What Is a Shadowban and Why Does It Matter? | Built In. https://builtin.com/marketing/shadowban

[111] Adi Robertson. 2019. TikTok prevented disabled users' videos from showing up in feeds. https://www.theverge.com/2019/12/2/20991843/tiktok-bytedance-platform-disabled-autism-lgbt-fat-user-algorithm-reach-limit

[112] Jeremiah Rodriguez. 2019. Instagram apologizes to pole dancers after hiding their posts. https://www.ctvnews.ca/sci-tech/instagram-apologizes-to-pole-dancers-after-hiding-their-posts-1.4537820 Section: Sci-Tech.

[113] Guy Rosen. 2019. Remove, Reduce, Inform: New Steps to Manage Problematic Content. https://about.fb.com/news/2019/04/remove-reduce-inform-new-steps/

[114] Riley Runnells. 2020. TikTok Denies Shadow Banning LGBTQ+ Hashtags. https://www.papermag.com/tiktok-lgbtq-shadow-banning-2647646779.html Section: LGBTQ.

[115] Koustuv Saha, Sindhu Kiranmai Ernala, Sarmistha Dutta, Eva Sharma, and Munmun De Choudhury. 2020. Understanding Moderation in Online Mental Health Communities. In *International Conference on Human-Computer Interaction*. 20.

[116] Salty. 2020. Shadowbanning is a Thing — and It's Hurting Trans and Disabled Advocates. https://saltyworld.net/shadowbanning-is-a-thing-and-its-hurting-trans-and-disabled-advocates/ Library Catalog: saltyworld.net Section: Algorithmic Bias.

[117] Salty. 2021. Exclusive: An Investigation into Algorithmic Bias in Content Policing on Instagram (PDF download) | Salty. https://saltyworld.net/algorithmicbiasreport-2/ Section: #MeToo.

[118] Joseph Seering, Tony Wang, Jina Yoon, and Geoff Kaufman. 2019. Moderator engagement and community development in the age of algorithms. *New Media & Society* (Jan. 2019), 1461444818821316. https://doi.org/10.1177/1461444818821316

[119] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday Algorithm Auditing: Understanding the Power of Everyday Users in Surfacing Harmful Algorithmic Behaviors. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 433:1–433:29. https://doi.org/10.1145/3479577

[120] Ellen Simpson, Andrew Hamann, and Bryan Semaan. 2022. How to Tame: LGBTQ+ Users' Domestication of TikTok. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (Jan. 2022), 22:1–22:27. https://doi.org/10.1145/3492841

[121] Ellen Simpson and Bryan Semaan. 2021. For You, or For"You"?: Everyday LGBTQ+ Encounters with TikTok. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan. 2021), 1–34. https://doi.org/10.1145/3432951

[122] Shakira Smith, Claire Fitzsimmons, and Oliver L. Haimson. 2021. Exclusive Report: Censorship of Marginalized Communities on Instagram, 2021) | Salty. https://saltyworld.net/product/exclusive-report-censorship-of-marginalized-communities-on-instagram-2021-pdf-download/

[123] Sanna Spišák, Elina Pirjatanniemi, Tommi Paalanen, Susanna Paasonen, and Maria Vihlman. 2021. Social Networking Sites' Gag Order: Commercial Content Moderation's Adverse Implications for Fundamental Sexual Rights and Wellbeing. *Social Media + Society* 7, 2 (April 2021), 20563051211024962. https://doi.org/10.1177/20563051211024962 Publisher: SAGE Publications Ltd.

[124] Liam Stack. 2018. What Is a 'Shadow Ban,' and Is Twitter Doing It to Republican Accounts? *The New York Times* (July 2018). https://www.nytimes.com/2018/07/26/us/politics/twitter-shadowbanning.html

[125] Lucy Suchman. 2011. Subject objects. *Feminist Theory* 12, 2 (Aug. 2011), 119–145. https://doi.org/10.1177/1464700111404205 00144 Publisher: SAGE Publications.

[126] Lucy Suchman and Lucy A. Suchman. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press.

[127] Nicolas P. Suzor, Sarah Myers West, Andrew Quodling, and Jillian York. 2019. What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation. *International Journal of Communication* 13, 0 (March 2019), 18. https://ijoc.org/index.php/ijoc/article/view/9736 Number: 0.

[128] Jeanna Sybert. 2021. The demise of #NSFW: Contested platform governance and Tumblr's 2018 adult content ban. *New Media & Society* (Feb. 2021), 1461444821996715. https://doi.org/10.1177/1461444821996715 Publisher: SAGE

Publications.

[129] Sam Tabahriti. 2022. Mark Zuckerberg says there is no ’shadow banning’ on Facebook but admits there are ’millions of mistakes’. https://www.businessinsider.com/mark-zuckerberg-no-shadow-ban-facebook-but-mistakes-are-made-2022-8

[130] Terry Tateossian. 2021. How to Fix Your Instagram Shadowban. https://www.entrepreneur.com/article/377057

[131] TeamYouTube. 2020. @Herclueless We don’t shadowban channels, but it’s possible the video was flagged by our systems as potentially violating guidelines. It may not show up in search, etc. before it’s reviewed. Since we have limited workforce due to COVID-19, reviews are taking longer: https://t.co/f25cOgmwRV. https://twitter.com/TeamYouTube/status/1319372516398452737

[132] TeamYouTube. 2020. @IslaDrummond Thanks for reaching out – YouTube doesn’t shadowban accounts. If you’re referring to live chats not working for owners and moderators, we’ve seen similar reports and are working on a fix. We’ll reach back out once we have more info to share. https://twitter.com/TeamYouTube/status/1340466309398720520

[133] TeamYouTube. 2020. @Simptress YouTube doesn’t shadowban channels. It’s possible the video was flagged by our systems as potentially violative & needs to be reviewed first before it shows up in search, etc. Note that reviews are taking longer since we have limited teams due to COVID-19: https://t.co/f25cOgmwRV. https://twitter.com/TeamYouTube/status/1319378407822589952

[134] Hibby Thach, Samuel Mayworm, Daniel Delmonaco, and Oliver Haimson. 2022. (In)visible moderation: A digital ethnography of marginalized users and content moderation on Twitch and Reddit. *New Media & Society* (July 2022), 146144482211098. https://doi.org/10.1177/14614448221109804

[135] TIkTok Newsroom. 2020. How TikTok recommends videos #ForYou. https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you

[136] Benjamin Toff and Rasmus Kleis Nielsen. 2018. “I Just Google It”: Folk Theories of Distributed Discovery. *Journal of Communication* 68, 3 (June 2018), 636–657. https://doi.org/10.1093/joc/jqy009

[137] Twitter. 2018. People are asking us if we shadow ban. We don’t. Read more to get all the facts. https://cards.twitter.com/cards/gsby/60efb. https://twitter.com/Twitter/status/1022658436704731136

[138] Kristen Vaccaro, Dylan Huang, Motahhare Eslami, Christian Sandvig, Kevin Hamilton, and Karrie Karahalios. 2018. The Illusion of Control: Placebo Effects of Control Settings. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI ’18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173590

[139] Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2020. “At the End of the Day Facebook Does What It Wants”: How Users Experience Contesting Algorithmic Content Moderation. *Proc. ACM Hum.-Comput. Interact* 4, CSCW2 (2020), 22.

[140] Julian Van Horne. 2020. Shadowbanning is a Thing — and It’s Hurting Trans and Disabled Advocates | Salty. https://www.saltysweethearts.com/shadowbanning-is-a-thing-and-its-hurting-trans-and-disabled-advocates/ Section: Algorithmic Bias.

[141] Julia Velkova and Anne Kaun. 2019. Algorithmic resistance: media practices and the politics of repair. *Information, Communication & Society* (Aug. 2019), 1–18. https://doi.org/10.1080/1369118X.2019.1657162

[142] waxpancake. 2009. What was the first website to hide troll’s activity to everyone but the troll himself? https://ask.metafilter.com/117775/What-was-the-first-website-to-hide-trolls-activity-to-everyone-but-the-troll-himself

[143] Riley Weeden. 2020. From shadowbanned to center stage: a student-led social media platform for those the algorithms leave behind. https://uofsdmedia.com/from-shadowbanned-to-center-stage-a-student-led-social-media-platform-for-those-the-algorithms-leave-behind/

[144] Lucas Wright. 2022. Automated Platform Governance Through Visibility and Scale: On the Transformational Power of AutoModerator. *Social Media + Society* 8, 1 (Jan. 2022), 20563051221077020. https://doi.org/10.1177/20563051221077020 Publisher: SAGE Publications Ltd.

[145] Mark Zuckerberg. 2018. A Blueprint for Content Governance and Enforcement. https://www.facebook.com/notes/751449002072082/

7 APPENDIX

A PLATFORM RESPONSES TO SHADOWBANNING

Table 4. Platforms’ Most Recent Known Public Responses to Shadowbanning and Deprioritization of Certain Social Media Content (at the time of this writing)

Platform	Platform Response to Shadowbanning	Source	Date

Twitter	“We do not shadow ban. You are always able to see the tweets from accounts you follow (although you may have to do more work to find them, like go directly to their profile). And we certainly don’t shadow ban based on political viewpoints or ideology.” [52]	Twitter Blogs blogpost about shadowbanning (or alleged lack thereof) from the company	July 26, 2018
Twitter	“People are asking us if we shadow ban. We don’t. Read more to get all the facts.” [137]	Official Twitter account in a tweet with a link to the company’s Shadowbanning blog post	July 26, 2018
Twitter	“We build our policies and rules with a principle of impartiality: objective criteria, rather than on the basis of bias, prejudice, or preferring the benefit to one person over another for improper reasons … In the spirit of accountability and transparency: recently we failed our intended impartiality. Our algorithms were unfairly filtering 600,000 accounts, including some members of Congress, from our search auto-complete and latest results. We fixed it.” [26]	Former Twitter CEO Jack Dorsey testifying before Congress about shadowbanning	September 5, 2018
Twitter	“Does Twitter shadow ban? Let’s discuss this one upfront. Simply put, we don’t shadow ban! Ever. We do rank Tweets to create a more relevant experience for you, however, and you’re always able to see Tweets from people you follow. Check out our company blog post for more details.” [23]	Answer to a question at the company’s Twitter Help Center	n.d.
Twitter	“Twitter is working on a software update that will show your true account status, so you know clearly if you’ve been shadowbanned, the reason why and how to appeal.” [100]	Twitter CEO Elon Musk announcing the development of a feature that can inform users if they are shadowbanned	Dec. 8, 2022
Instagram	“Today we discussed how Instagram is working to ensure that the content we recommend to people is both safe and appropriate for the community. We have begun reducing the spread of posts that are inappropriate but do not go against Instagram’s Community Guidelines, limiting those types of posts from being recommended on our Explore and hashtag pages. For example, a sexually suggestive post will still appear in Feed if you follow the account that posts it, but this type of content may not appear for the broader community in Explore or hashtag pages.” [113]	Meta News Release	April 10, 2019

Instagram	“Shadowbanning is not a thing. Someone follows you on Instagram, your photos and videos can show up in their feed if they keep using their feed and being in Explore is not guaranteed for anyone. Sometimes you get lucky. Sometimes you won’t.” [88]	Adam Mosseri, Instagram CEO, on an Instagram Live conversation when asked, “Shadowbanning. It’s not a thing, right?	Feb. 22, 2020
Instagram	“While some posts on Instagram may not go against our Community Guidelines, they might not be appropriate for our global community, and we’ll limit those types of posts from being recommended on Explore and hashtag pages. For example, a sexually suggestive post will still appear in Feed if you follow the account that posts it, but this type of content may not appear for the broader community in Explore and hashtag pages.” [74]	Current Instagram policy on certain posts’ absence from Explore pages as of the writing of this article	2021
Instagram	“People often accuse us of “shadowbanning” or silencing them. It’s a broad term that people use to describe many different experiences they have on Instagram. We recognize that we haven’t always done enough to explain why we take down content when we do, what is recommendable and what isn’t, and how Instagram works more broadly. As a result, we understand people are inevitably going to come to their own conclusions about why something happened, and that those conclusions may leave people feeling confused or victimized. That’s never our intention, and we’re working hard on improvements here. We also manage millions of reports a day, which means making a mistake on even a small percentage of those reports affects thousands of people. We also hear that people consider their posts getting fewer likes or comments as a form of ’shadowbanning.’ We can’t promise you that you’ll consistently reach the same amount of people when you post. The truth is most of your followers won’t see what you share, because most look at less than half of their Feed. But we can be more transparent about why we take things down when we do, work to make fewer mistakes – and fix them quickly when we do – and better explain how our systems work. We’re developing better in-app notifications so people know in the moment why, for instance, their post was taken down, and exploring ways to let people know when what they post goes against our Recommendations Guidelines.” [98]	Adam Mosseri, Instagram CEO, on Shadowbanning in a blog post	2021

Instagram	“A number of hashtags, including #poledancenation and #polemaniabr, were blocked in error and have now been restored. We apologise for the mistake. Over a billion people use Instagram every month, and operating at that size means mistakes are made – it is never our intention to silence members of our community.” [4][112]	Facebook spokesperson in an email to Carolina at the blog “Blogger on Pole” in regard to hiding posts containing hashtags related to pole dancing on Instagram	July 31, 2019
Instagram	“We understand users have experienced issues with our hashtag search that caused posts to not be surfaced. We are continuously working on improvements to our system with the resources available. When developing content, we recommend focusing on your business objective or goal rather than hashtags. Having a growth strategy that targets the right audience is essential to success on Instagram. Good content on Instagram is simply good creative. And it follows the same three creative principles you’d apply to any marketing channel: - Have a distinct visual presence: Include your logo, an iconic brand element, a brand color or even a product you’re known for to make your content distinct and easily recognizable for the community. - Be a storyteller: Tell a story that supports your business goal. Whether you want to raise awareness or increase sales of a specific product, make sure the imagery and copy the latter up to your main goal. - Put thought into your creative: Be well crafted to stand out. This doesn’t mean you need to build additional content for Instagram. It just means you need to put as much love and care into the content to inspire as you do in your business. We truly appreciate your understanding and patience in this matter.” [49](Instagram for Business, 2017)	Post by the Instagram for Business account on Facebook	February 28, 2017
Instagram	“We know that some people are experiencing issues uploading and viewing stories. This is a widespread global technical issue not related to any particular topic and we’re fixing it right now. We’ll provide an update as soon as we can.” [75]	Tweet from the Instagram Comms official account in response to suppression of content showing the Israeli government’s continued violence against Palestinians	May 6, 2021

Instagram	"Today we are adding new transparency tools so that you can see whether or not your photos and videos are recommended in places like "Explore." ...What we've added is [a feature that shows] whether or not your account can be recommended in places like "Explore." If you've posted things that violate our "Recommendation Guidelines"... you can end up in a state where your content won't be recommended. You can edit that content, delete that content, or appeal if you disagree." [99]	Instagram post from Adam Mosseri, Instagram CEO, introducing a feature allowing Instagram users to see whether Instagram has flagged their content for violating "Recommendation Guidelines"	Dec. 7, 2022
YouTube	"YouTube doesn't shadowban channels. It's possible the video was flagged by our systems as potentially violative and needs to be reviewed first before it shows up in search, etc. Note that reviews are taking longer since we have limited teams due to COVID-19." [133]	@TeamYouTube Twitter Account in response to Tweet Thread about shadowbanning	Oct. 22, 2020
YouTube	"Thanks for reaching out – YouTube doesn't shadowban accounts. If you're referring to live chats not working for owners and moderators, we've seen similar reports and are working on a fix. We'll reach back out once we have more info to share." [131]	@TeamYouTube Twitter Account in response to Tweet Thread about shadowbanning	Dec. 19, 2020
YouTube	"We don't shadowban channels, but it's possible the video was flagged by our systems as potentially violating guidelines. It may not show up in search, etc. before it's reviewed. Since we have limited workforce due to COVID-19, reviews are taking longer." [132]	@TeamYouTube Twitter Account in response to Tweet Thread about shadowbanning	Oct. 22, 2020
Facebook	"We train AI systems to detect borderline content so we can distribute that content less." [145]	Mark Zuckerberg discussing deprioritization of borderline content in "A Blueprint for Content Governance and Enforcement"	Nov. 2018
Facebook	"We also deploy several approaches to reduce the prevalence of violating content, such as: removing accounts, Pages, Groups and events for violating our Community Standards or Guidelines; filtering problematic Groups, Pages and content from recommendations across our services; or reducing the distribution of likely violating content or content borderline to our Community Standards." [35]	Statement from Meta about Facebook reducing distribution of "borderline" content	Nov. 9, 2021

Facebook	“There is no policy that is ’shadow banning,’ so I think it’s sort of a slang term... that maybe refers to some of the demotions [of posts] that we’re talking about... [If a post] is marked as false by a fact-checker, it will get somewhat less shown... but if there’s some history within a page, then there can be some kind of broader policy that applies.” [129]	*Business Insider*’s summary of Mark Zuckerberg’s statements made on the *Joe Rogan Podcast* regarding the deprioritization of content on Facebook	Aug. 28, 2022
TikTok	“Our TikTok Creator Marketplace protections, which flag phrases typically associated with hate speech, were erroneously set to flag phrases without respect to word order. We recognize and apologize for how frustrating this was to experience, and our team is working quickly to fix this significant error. To be clear, Black Lives Matter does not violate our policies and currently has over 27 billion views on our platform” [29]	TikTok spokesperson to Insider Magazine in response to claims of suppressing Black Lives Matter content	July 8, 2021
TikTok	“This approach was never intended to be a long-term solution and although we had a good intention, we realised that it was not the right approach” [85]	TikTok spokesperson to *Netzpolitik* in response to leaked documents claiming that TikTok moderators were instructed to suppress the content of disabled, queer, and/or fat users	Dec. 2, 2019
TikTok	“Early on, in response to an increase in bullying on the app, we implemented a blunt and temporary policy. While the intention was good, the approach was wrong and we have long since changed the earlier policy in favor of more nuanced anti-bullying policies and in-app protections.” [111]	TikTok spokesperson to *The Verge* in response to leaked documents claiming that TikTok moderators were instructed to suppress the content of disabled, queer, and/or fat users	Dec. 9, 2019

TikTok	"On May 19, Black creators and allies took an important stance in changing their profile pictures and connecting on the platform to speak out against how they feel the Black community has been marginalized on TikTok. And at the height of a raw and painful time, last week a technical glitch made it temporarily appear as if posts uploaded using #BlackLivesMatter and #GeorgeFloyd would receive 0 views. This was a display issue only that widely affected hashtags at large, and powerful videos with the #BlackLivesMatter hashtag continued to be uploaded, viewed, and engaged with – in fact, videos with these hashtags have currently generated well over 2 billion views, which is a testament to their importance to and resonance among our community. Nevertheless, we understand that many assumed this bug to be an intentional act to suppress the experiences and invalidate the emotions felt by the Black community. And we know we have work to do to regain and repair that trust." [106]	Statement from TikTok leadership in response to claims of suppression of content using #BlackLivesMatter and #GeorgeFloyd	June 1, 2020
TikTok	"Since users spend most of their time exploring their For You feeds, last week we also took an important step in providing insight into our recommendation system to help users understand their options for shaping their unique experience. We want to be open about the inherent challenges recommendation systems face and how we work to protect against bias. Our teams are intent on developing ethical machine learning processes that reflect inclusivity and diversity, but we know there's work to be done in this area and we are committed to further research and investment toward that goal." [107]	TikTok Progress report from the company in relation to their statement about the suppression of Black creators	June 24, 2020
TikTok	"One of the inherent challenges with recommendation engines is that they can inadvertently limit your experience – what is sometimes referred to as a "filter bubble." By optimizing for personalization and relevance, there is a risk of presenting an increasingly homogenous stream of videos. This is a concern we take seriously as we maintain our recommendation system." [135]	Statement from TikTok about how their algorithmic For You Page works and can prioritize or suppress certain content on one's individual page	June 18, 2020

B SURVEY AND INTERVIEW INSTRUMENTS

This appendix only includes parts of the survey and interview that were included in this paper's analysis. Several survey questions were adapted from OnlineCensorship.org [105] and Myers West [101].

B.1 Shadowbanning survey questions

(1) Within the last year, have you personally experienced shadowbanning on social media site? [Yes; No; I'm not sure]
(2) On which social media platform(s) did the shadowbanning occur? [list of social media sites, multiple options possible]

(3) Please describe your experience with shadowbanning. [open-ended]
(4) Why do you think the shadowban happened? [open-ended]

B.2 Shadowbanning interview questions

Interviews were semi-structured, so in addition to these questions, we asked follow-up questions and focused on aspects of shadowbanning experiences most salient to participants.

(1) Have you ever heard of shadowbanning?
(2) Is this something you have experienced or seen on social media?
(3) Can you tell us more about this experience? How did you know that shadowbanning was happening?