"How advertiser-friendly is my video?": YouTuber's Socioeconomic Interactions with Algorithmic Content Moderation

RENKAI MA, Pennsylvania State University, USA
YUBO KOU, Pennsylvania State University, USA

To manage user-generated harmful video content, YouTube relies on AI algorithms (e.g., machine learning) in content moderation and follows a retributive justice logic to punish convicted YouTubers through demonetization, a penalty that limits or deprives them of advertisements (ads), reducing their future ad income. Moderation research is burgeoning in CSCW, but relatively little attention has been paid to the socioeconomic implications of YouTube's algorithmic moderation. Drawing from the lens of algorithmic labor, we describe how algorithmic moderation shapes YouTubers' labor conditions through algorithmic opacity and precarity. YouTubers coped with such challenges from algorithmic moderation by sharing and applying practical knowledge they learned about moderation algorithms. By analyzing video content creation as algorithmic labor, we unpack the socioeconomic implications of algorithmic moderation and point to necessary post-punishment support as a form of restorative justice. Lastly, we put forward design considerations for algorithmic moderation systems.

CCS Concepts: • **Human-centered computing → Collaborative and social computing → Empirical studies in collaborative and social computing**

KEYWORDS: Content moderation; algorithmic moderation; YouTube moderation; socioeconomics; algorithmic labor; YouTuber

ACM Reference format:

Renkai Ma and Yubo Kou. 2021. "How advertiser-friendly is my video?": YouTuber's Socioeconomic Interactions with Algorithmic Content Moderation. In *PACM on Human Computer Interaction,* Vol. 5, CSCW2, Article 429, October 2021. ACM, New York, NY, USA. 25 pages. https://doi.org/10.1145/3479573

1 INTRODUCTION

YouTube has become the largest video-sharing platform. "Broadcast Yourself," YouTube's slogan, implies this platform is primarily for ordinary people who want to create and share videos, and two billion registered users[1] worldwide today can post video content or consume others' content. Those video content creators (or YouTubers) can also join the YouTube Partner

This work is partially supported by the National Science Foundation, under grant no. 2006854.

Author's addresses: Renkai Ma (renkai@psu.edu) and Yubo Kou (yubokou@psu.edu), College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, 16802, USA

[1] https://www.omnicoreagency.com/youtube-statistics/

2021 2573-0142/2021/10 – Art429 $15.00

https://doi.org/10.1145/3479573

Program (YPP)[2] to earn advertisement (ad) revenue, which refers to 'monetization' where video creation and sharing become profits[3] [56,73]. Nowadays, more YouTubers' livelihoods rely upon the business of making videos on YouTube [2,15,17,50]. Thus, while content creation on commonly examined platforms like Reddit and Twitter is usually framed along the line of free expression, content creation on YouTube is distinct in how it is weaved into the platform economy and manifests as a form of digital labor [73,82].

But not all video content is advertiser-friendly, and not all labor is deemed worthy of compensation. Particularly, harmful content such as hate speech and racism [63] is detrimental to both YouTube's business model and its user community. Like other social media sites, YouTube also wrestles with the grim challenge of content moderation [39]. Content moderation refers to online governance mechanisms that regulate inappropriate content such as hate speech, harassment, and violence to facilitate cooperation and prevent abuse [44]. CSCW and HCI researchers have focused on sociotechnical aspects of content moderation, such as sociotechnical mechanisms of moderation or social practices of moderators [47,48,95,96]. For example, on Reddit, voluntary human workers in their subreddits can manually moderate or utilize machine learning algorithms to regulate harmful content [47]. Twitter exclusively relies on voluntary users to report harmful tweets, and then algorithms handle them behind the scenes [20].

However, what is less discussed in the literature is the socioeconomic implication of content moderation: YouTube's content moderation economically punishes YouTubers, its laborers. Once a YouTuber is determined to have created harmful content, YouTube would demonetize[4] their user accounts or videos, eventually denying them from earning more future ad revenue through limited or no ads placed on videos. Currently, little attention has been paid to understand YouTubers' socioeconomic interactions with algorithmic moderation or, in other words, how YouTubers interact with the socioeconomic punishments (i.e., demonetization) caused by YouTube moderation.

To answer this question, we gathered and analyzed discussion data from the 'r/youtube'[5] subreddit, nearly the largest YouTube-related online community today. Utilizing an inductive thematic analysis [59], we identified how YouTubers perceived, experienced, and reacted to algorithmic moderation punishments. We found that opacity of algorithmic punishments existed in multiple layers, and such opacity led YouTuber's video creation work to be precarious. Also, YouTubers sought to cope with moderation punishments, in a reflexive manner, by gradually gaining and applying practical knowledge of algorithms. Drawing from the lens of algorithmic labor [79], a form of digital labor associated with sophisticated algorithmic systems, we discuss a socioeconomic understanding of algorithmic moderation on YouTube and how YouTubers shared and received support to speculate, make sense of, and reflect on algorithmic penalties, informing their behaviors of repairing and avoiding future punishments. We then showed how peer and platform support could serve as a restorative

[2] https://support.google.com/youtube/answer/72851?hl=en

[3] https://support.google.com/youtube/answer/72857?hl=en

[4] Demonetization is an idiomatic term and a moderation outcome describing the decrease or deprivation of future ad income due to various YouTube moderation such as limited advertisements (ads), no ads, copyright infringement, age-restriction, or other moderation decisions. Many YouTubers frequently refer to demonetization exclusively as one moderation decision, 'limited ads,' where YouTube deem that most advertisers are not willing to place ads on those videos. However, as we stated, multiple moderation decisions could cause demonetization penalty/outcome.

[5] https://www.reddit.com/r/youtube/

justice means. Ultimately, we put forward design considerations for algorithmic moderation systems.

This research contributes to the CSCW literature by 1) initially investigating the user experience of content moderation from the angle of video content creators, 2) providing empirical evidence and conceptual understandings of how YouTubers interact with algorithmic content moderation, 3) implicating design considerations of algorithmic systems to be more transparent and accountable in content moderation, and 4) connecting CSCW research on social media's content moderation with socioeconomic perspectives, beyond the often discussed sociotechnical aspects.

2 BACKGROUND: Socioeconomic Content Creation, Content Policy on YouTube

YouTube is unique in its socioeconomic features of video content creation. Socioeconomics here denotes that economic activities mutually affect social behaviors [21,45]. By joining the YouTube Partner Program (YPP), YouTubers can earn revenue from the advertisements (ads) inserted within their videos. Ad revenue is calculated based on the viewing quantity of videos; YouTubers' videos can catch more new audiences' attention when YouTubers directly and effusively interact with audiences [7], indicating more social engagement behaviors around videos could lead to higher ad revenue. By interacting more with audiences, YouTubers can be aware of what videos could be more lucrative by suiting viewer tastes, forming the socioeconomic content creation on YouTube.

YouTubers might create harmful content, which is a consistent issue for YouTube. YouTube comprehensively classified multiple harmful content topics in its AdSense Google publisher policy[6], community guidelines[7], and advertiser-friendly content guidelines (ACG)[8]. All these content policies explicitly prohibit videos containing harmful content such as harassment [103], sexist hate speech [25,26], sexually suggestive materials, and terrorism [65,84]. One example of harmful content is that the YouTuber Logan Paul uploaded a video titled "We found a dead body in the Japanese Suicide Forest" in 2018 and showed graphic footage of a suicided victim's body in Japan's Aokigahara forest [54]. Another example is that Carlos Maza, a Vox media journalist, demanded YouTube to punish the YouTuber Steven Crowder in 2019 because his videos contained more than two years of harassment and homophobic hate speech on Carlos's sexual orientation and ethnicity [90].

To alleviate or prevent the negative effects of YouTubers' harmful content, YouTube consistently tightens content policies. YouTube frequently modifies its ACG to protect its advertisers. For instance, "YouTube Adpocalypse," an internet slang describing a phenomenon where advertisers stop their advertisements (ad) on YouTube because specific bad content harms their brand image, triggers YouTube to introduce stricter content policies[9]. Similarly, YouTube updates content policies to resonate with governmental requirements, indirectly benefiting advertisers. For instance, in September 2019, "the US's Federal Trade Commission issued a $170 million fine against Google for alleged violations of the children's online privacy protection act (COPPA)" [3]. YouTube thus released a new content policy requiring all YouTubers worldwide to set targeted audiences for their channels and videos between tags of

[6] https://support.google.com/adsense/answer/9335564
[7] https://www.youtube.com/howyoutubeworks/policies/community-guidelines/#developing-policies
[8] https://support.google.com/youtube/answer/6162278?hl=en
[9] https://youtube.fandom.com/wiki/YouTube_Adpocalypse

"made for kids" and "not made for kids."[10] Collectively, improving content policy can help YouTube maintain its social and business image among audiences (i.e., advertisers' potential customers [106]) and advertisers, ensuring future videos on YouTube to be acceptable and advertiser-friendly.

3 RELATED WORK

We situate our study in prior scholarship regarding algorithmic content moderation on social media, debates of its transparency, user behaviors after moderation punishments, and algorithmic labor of video content creation.

3.1 Algorithmic Content Moderation and Its Socioeconomic Effects

Content moderation on social media usually needs to balance cost and efficiency [44]. Given the sheer and increasing volume of user-generated content on social media, there are not enough human moderators available to scrutinize each new piece of content [20,47]. Also, manually moderating content is generally time-consuming and impossible for practice [89]. Thus, many social media platforms have turned to AI algorithms (e.g., machine learning) [10,43,62] to automate content moderation, at least partially. For example, Facebook uses algorithmic tools to flag group-join requests from identified spam users automatically [70]. Twitch, a live streaming platform, uses automatic moderators to regulate content in the chatrooms between live streamers and audiences [86].

Algorithmic content moderation takes multiple sociotechnical forms on social media. In the aspect of technical design, social media such as YouTube, Facebook, Reddit, and Twitter relies on different machine learning algorithms to regulate users' content [4,43,47,62]. For example, platforms frequently use natural language processing, speech recognition, or sentiment analysis to recognize harmful content and fake news [41]. Regarding moderation's power allocations, human moderators can play an important role: voluntary users or commercially trained flaggers employed by social media companies manually flag or review users' content [20].

Content moderation affects users primarily through punishments, ranging from content removal to user account suspension [35]. For instance, social media can also shadow-ban accounts, which means users can still post content without recognizing the punishments, but their content will be invisible to other users until moderators approve [14,19,66]. Social media like Instagram or Tumblr can also ban hashtags and the associated discussions, where punishments are executed on a platform level, affecting all platform users. Moderation punishments can influence users' future behaviors in significant ways [66] and have been criticized for heavily censoring free expression [92].

What is distinctive about content moderation on YouTube is that its algorithmic punishments have socioeconomic implications. Once YouTubers violate content policies, they would experience not only sociotechnical forms of moderation similar to other social media users but also a socioeconomic punishment: demonetization [13], referring to deducting or deprive the future ad revenue of a video or YouTuber channel. Given the concept of socioeconomics that economic activities mutually affect social actions [21,45], demonetization impacts might motivate punished YouTubers to adjust their future behaviors to weaken such demonetization effects [13] for steady ad income.

[10] https://support.google.com/youtube/answer/9527654

While previous moderation literature has explored the sociotechnical implications of moderation on platform users' experiences, relatively little attention has been paid to the intersections of algorithmic moderation and socioeconomic impacts. This study seeks to address this gap at the angle of content creators.

3.2 Algorithmic Transparency and Post-Punishment Behavior

Social media platforms have been criticized for the limited transparency of their algorithmic content moderation [33,53,74]. Researchers pointed out that platforms moderate users and their content in a murky way without enough explanations [39]. For instance, Juneja et al. [52] discovered that Reddit's moderation violated Santa Clara Principles (SCP) of Transparency and Accountability[11] in aspects including an absence of explanations for removed content and ambiguous removal led by implicit community norms instead of content policies.

Moderation explanations are deemed important in helping the user understand moderation mechanisms. Moderation decisions are generally accompanied by short, formal, and ambiguous explanations [39]. Sometimes, online communities' content policies are also vaguely worded. This less transparent moderation could make social media users feel unfair and frustrated [46], lead them to generate folk theories for their future online operations [32], or develop biased beliefs to explain moderation decisions [66]. Recent research has pointed out that the provision of explanations helped users clearly understand content policies. One example is that Jhaver et al. [48] found that when provided moderation explanations, Reddit users were then intended to learn explicit content guidelines in specific subreddits. Also, Kou and Gui [57] further pointed out that explanations should include community context (e.g., shared values, knowledge, and community norms) to make users understand how algorithms could work better for end-users.

A good explanation should be generated by an explainable decision-making process of algorithms. One of the challenges in HCI research and practice grounded by Shneiderman et al. is that novel systems should allow users to understand invisible algorithmic processes to better control their future actions [88]. Resonating with this call, various interests have uncovered the importance of human-AI collaboration to improve the trust of algorithmic decision-making [97]. For example, Wang et al. tested how different explanation types produced by explainable AI (XAI) systems support users' reasoning to improve trust for the system and mitigate users' cognitive bias [98]. Similarly, moderation systems that were found to provide explanations of appeal can improve users' perceptions of fairness, trust, and transparency [95].

Parallelly, more studies have started to investigate how social media users cope with moderation after experiencing moderation punishments. For example, Jhaver et al. [48] found on Redditt, more explanations provided in algorithmic moderation were associated with more users' content-generating behaviors complying with content policies. Cobbe [18] theoretically summarized two strategies of successfully resisting algorithmic content moderation on social media: everyday resistance and organized resistance [83]. Everyday resistance refers to small-scale and relatively safe circumventing activities. Like what Gerrard [38] unearthed, punished users could evade platform policies to develop alternative hashtags on Instagram and Tumblr. Besides, organized resistance means that collective behaviors undermine the power of social media's algorithmic content moderation. For instance, Chancellor et al. [16] found that users

[11] https://santaclaraprinciples.org/

could collectively form closed societal groups to avoid future moderation attention on Instagram.

Particularly, socioeconomic punishments from YouTube's algorithmic moderation have been reported to be opaque. On the one hand, how YouTube decides a video as "unsuitable for advertisers" does not necessarily align with advertisers' attitudes [40]; at the same time, YouTube demonetizes videos that contain sensitive topics (e.g., subjects related to the war or natural disasters), forming a disincentive for news disseminating among audiences [22]. On the other hand, little is known about YouTube's moderation algorithms. YouTube's algorithms are the "black-box" where people hardly know how demonetization decisions are made [75]. For instance, YouTube might unfairly demonetize videos in different languages [72]. Researchers and journalists have also accumulated ample evidence of how content produced by minority groups is disproportionately demonetized without sufficient explanations [34,100].

So far, little is known as to how YouTubers' subjective experiences with their socioeconomic punishments. One exception is that Caplan [13] primarily investigated YouTube videos to understand YouTube's tiered governance where YouTube was deemed to unfairly demonetize and disproportionally distribute resources between small and large YouTubers (i.e., with large subscription number). However, by standing at the (video) content creator's perspective, there is still a lack of systematic investigations on how YouTubers perceive and learn from the algorithmic moderation's decisions as well as how they handle the socioeconomic punishment, namely demonetization. This study aims to fill this research gap.

3.3 Video Content Creation as Algorithmic Labor

Prior literature on social media moderation has oftentimes framed users' content creation activity primarily as a form of expression. For example, West's survey study of users who experienced content moderation drew primarily from "the lens of free expression" [66] to make a nuanced case for how we should understand content moderation's other implications, such as users' affective relationships with platforms, users' agency in interacting with platforms, and the educational potential of future moderation systems. Chancellor et al.'s linguistic analysis of pro-eating disorder content also considered user-generated content as speech and developed a large-scale quantitative analysis of the content's lexical variation [16]. However, on YouTube, users do not just make speeches through their video creation. Their video creation is a form of digital labor [73]: Even if some of them do not intend to profit from it, their work is still organically incorporated into the platform economy of YouTube — they provide immaterial labor for YouTube.

Researchers have long investigated how algorithms mediate labor. For example, Raval and Dourish's study of ridesharing workers showed how Uber and Lyft drivers must strive for a better rating determined by the ridesharing algorithms [76]. Through Uber's central algorithmic system, power and information asymmetries arise as they surveil and shape drivers' behaviors [79]. On YouTube, Youtubers are also engaged and enmeshed in webs of algorithms. YouTubers need to acquire and benefit from knowledge about how video recommendation algorithms work [4,8,9,104]. Digital influencers, a category of algorithmic laborers, have to figure how the algorithmic rules on Instagram in order to enhance their visibility and subsequently profit [19].

The algorithmic labor literature has explored how the primary algorithmic mechanisms, such as work allocation and recommendation, shape content creators' labor conditions. However, given that social media platforms have historically placed moderation concerns at the periphery of their business logic [39], it is somewhat unsurprising that little attention has been paid to the

labor implications of moderation algorithms. Thus, in this paper, we use the lens of algorithmic labor to examine YouTubers' video creation work, with a focus on how their labor is intertwined with moderation algorithms.

4 METHODS

4.1 Data Collection

In this study, we used discussion data of the 'r/youtube' subreddit. It is the largest online community having more members than any other YouTube-related forums such as yttalk[12]. YouTubers' online discussions could contain abundant experiences they share naturally and directly. Besides, online discussions could uncover how YouTubers interact with algorithmic moderation collectively and show how they communicate and collaborate.

We iteratively fetched relevant threads discussing YouTube moderation from 'r/youtube.' We ran the package 'RedditExtractoR' [77] on R 4.0.4 to fetch the threads by relevant keywords. Upon Reddit's API, this package allowed us to filter out all historical threads having specific keywords in either content, user comments, or titles. Therefore, we first generated a preliminary list of keywords. We synthesized the keywords related to YouTube's content moderation from the literature discussing social media content moderation (e.g., [20,36,47,52,66,95]) and YouTube content moderation (e.g., [4,13]), as well as relevant media reports (e.g., [1,67]). We generated the initial keywords: {moderate, censor, ban, delete, violate, suspend, demonetize, remove, shadowban, terminate, algorithm, transparent, ad-friendly (advertiser friendly), flag, appeal}. We then searched by all forms of keywords (e.g., for the word "demonetize," it corresponds to demonetize, demonetized, and demonetization) to ensure search results' completeness. After removing duplicates by each unique combination of comment and comment date, we fetched an initial dataset containing 2,779 threads associated with 60,310 individual comments.

Second, we consistently searched relevant threads to make sure the dataset was as comprehensive as possible. We randomly read 50 threads to generate complementary keywords from this initial dataset. The purpose of this action is to include more contextual and spoken keywords that YouTubers often used to describe content moderation. Those additional keywords, {bot, restore, scan, re-scan (rescan), yellow monetization, swear (swearing)[13]}, are for the second-round search. We continued this process to add additional keywords to iteratively search more threads until the dataset became saturated [11], indicating there would be no new information provided. This iterative data collection process aimed to tolerate a false-positive rate where any data points mistakenly identified can be collected to generate a more informative dataset.

Third, we did two steps of data preprocessing operations. We calculated each comment's length and then deleted the comments with a length of fewer than 50 characters because we found these comments conveyed limited meanings, such as short replies to other users, asking questions, or posting with links. For example, these posts could be agreeing with others by commenting, *"Yeah, seems like YouTube is getting worse now..."* and *"Ahhhh, ok, thanks a lot for the clarification!!! :)"*. We then repetitively retrieved all threads by keywords and reviewed each thread with its posted text to determine whether it was related to content moderation;

[12] https://yttalk.com/

[13] YouTubers use curse words such as "damn," and "hell" in their videos.

unrelated threads and their associated comments were removed. For example, one of the irrelevant posts discussed the rumor that YouTube plans to remove the 'dislike' button.

The final dataset contained 3,086 threads and the associated 36,279 comments. It had variables including 'comment date,' 'the number of comments in the thread,' 'comment text,' 'title of the thread,' and 'the posted text of thread.' In detail, the comment date ranged from June 8, 2012, to December 23, 2020, and the average number of comments of each thread was around 21.82.

4.2 Data Analysis

We used an inductive thematic analysis [59] to probe the research question. The data analysis was processed by the same researchers who collected the data. We started to consistently read from the top 100 most discussed threads and their comments to the least commentated ones (i.e., threads that have no comments) and ran the 'open coding' to generate first-level codes. In this process, we focused on the discussion data explicitly disclosing posters' identities as either YouTubers or audiences. We labeled each code to be associated with correspondingly sentences or paragraphs to classify findings from the analysis process. Simultaneously, the authors ran 'axial coding' to collaboratively connect codes to build up higher-level concepts, where we resonated with prior studies if they were related. In this process, we also allowed new codes emerging from consistently reviewing data to genialize new themes. We ultimately finalized the analysis by both exerting 'selective coding' to connect higher concepts and thus acquiring a sound thematic closed loop, providing satisfactory and informative concepts for the research question.

We integrated and distilled into three high-level themes for the research question in the repetitive rounds of coding. These themes included "Being Confused in Algorithmic Opacity" (discussed in Section 5.1), "Managing Algorithmic Precarity" (Section 5.2), and "Learning and Applying Algorithmic Know-How" (Section 5.3).

4.3 Ethics Statement

We believe our study utilizing discussion data on the 'r/youtube' subreddit imposed minimal ethical risk to YouTubers and Reddit users. After our university's Institutional Review Board (IRB) approved this study, we performed data collection and analysis. We removed user information such as usernames and the URLs of posts, so the dataset did not contain any personally identifiable information. When presenting the findings, we paraphrased each quote to decrease its searchability on Redditt and used the singular "they" pronoun to interpret data, further assuring our dataset's anonymity.

5 FINDINGS: How do YouTubers Interact with Algorithmic Moderation?

By examining YouTuber's discussions regarding socioeconomic punishments, we uncovered three primary themes. YouTubers were confused to varied degrees due to the opacity of punishments. Such algorithmic opacity led their video creation labor on YouTube to be precarious. YouTubers then sought to handle past punishments and avoid future moderation risks, in a reflexive manner, by gaining and applying algorithmic know-how.

5.1 Being Confused in Algorithmic Opacity

Algorithmic opacity in moderation refers to the situations where YouTubers who experienced algorithmic punishments felt confused and had no clues of how algorithms made decisions. Even if the adjudication of a moderation case is clear-cut to most people, the YouTuber who received the penalty could experience it differently and feel confused. We found that YouTubers experienced algorithmic opacity at multiple layers. Moderation decisions could sometimes puzzle the ordinary audience. One viewer wrote:

> I do not post videos. I do not comment on anything. I watch makeup videos, clothing hauls, music, and random video game. There are zero reasons I should have my account suspended. I would love to know because I literally just watch videos.

The viewer perceived opaque moderation. Although the viewer believed they used YouTube in benevolent ways, the platform's moderation made a seemingly opposite decision. If a certain type of content triggered the moderation action, they would be able to reason about the moderation rationale. But in this case, the viewer could not associate any prior behavior with the punishment. While it is possible that this viewer here might not have fully disclosed the information they received, their confusion on algorithmic decisions was still a valid human experience.

Many YouTubers reported that they experienced algorithmic penalties without promised warnings beforehand. For example, one YouTuber wrote:

> My channel has been automatically suspended without any warning. I received the first email with no explanation. Then after I sent another email, they said it was based on community guidelines. I reviewed the guidelines, and I did not violate any of those things. If I did, I received no strikes or warnings like they promised to do.

In this quote, 'strike' refers to a violation notification from YouTube. The YouTuber here showed three occasions of experiencing the opacity of YouTube moderation. First, the YouTuber received nothing even though the policies said there would be warnings (i.e., strikes) beforehand. Second, insufficient or unconvincing explanations were provided after account suspension punishment. Third, the explanations provided by the YouTube service team were perceived as unconvincing to the YouTuber. This example resonated with how AI-based systems' explanations might fail to support user's reasoning [98].

When experiencing algorithmic moderation perceived as inaccurate, YouTubers suspected that algorithmic mechanisms needed more training data (i.e., video content). Still, because of moderation systems' complexity and insufficient explanation provision, YouTubers would perceive algorithms as opaque to understand, which resonates with the opacity of algorithmic moderation that researchers found on Reddit [47,52]. But what intensified such opacity was the moderation irregularity YouTubers experienced. One YouTuber wrote:

> I did have one of my flagged videos suddenly go back to being monetized last week, but within an hour or two, it was back to being flagged again. So, I am not sure if that was a sign of the bot learning and then re-flagging for some other reason or if it was just a slight glitch with the site.

The flag here refers to the demonetization penalty, 'limited ads.'[14] In the above case, the YouTuber might expect the algorithms to be consistent with the time dimension where each video was supposed to experience a one-time algorithmic inspection. They did not appeal to tell

[14] https://support.google.com/youtube/answer/9269824?hl=en

the bot if its demonetization penalties were perceived wrong. Instead, the bot here imposed penalties and corrected mistakes irregularly. This irregularity caused the YouTuber to speculate about how algorithms worked behind automated penalties. This example showed the YouTuber believed that bots needed time to learn from video data but still expressed their confusion on algorithmic decision-making over time.

When algorithmic punishments were accompanied with explanations, they often appeared insufficient to YouTubers. One YouTuber quoted from an email that YouTube sent to them and interpreted:

> Your account has been terminated as a result of repeated abusive, hateful, and/or harassing comments that violate our community guidelines." Stupidly vague. It seems everyone gets the same message. They do not tell you which post or posts you made that were reported and how they violated community guidelines. At the very least, they could screenshot the post that was reported. Nope. Nothing.

This YouTuber described their experience of account suspension for violating community guidelines without being informed with detailed reasons. They also observed a general phenomenon that YouTubers typically received similarly phrased moderation reasons. As Stohl et al. [91] suggested, moderation systems should offer the appropriate degree of information. Here, the YouTube platform did not specify which part of the video violated content policies. Instead, its explanations were generic and not situated in specific cases. These frequent complaints manifested the low confidence of explanation caused the YouTuber's negative sentiments toward the opacity of YouTube moderation. While the YouTuber here might hide critical information or keywords in the moderation explanation, the moderation algorithms seemly failed to reach the goals of human interpretability [27], generating confusion.

When YouTubers appealed past penalties, they often received a response inconsistent with their expected repairing process for moderation punishments. One YouTuber wrote:

> My channel with 20,000 subs being suspended. I have since filed an appeal but received an automated email response stating that they have decided to uphold the ban after only reviewing for 3 minutes.

Here, 'appeal'[15] originally refers to applying for a human review process to double-check the said violation. The YouTuber here claimed that they received the email that seemed to be an automated response. The human review, as the YouTuber thought, would take enough time to process. However, the appeal result was delivered within a short time, which conflicted with the YouTuber's shared sense of understanding of the appeal process. The "three minutes" here might be a sign that YouTube might have utilized algorithms for the appeal process. While the YouTuber here did not disclose the email details, this case showed the confusion and dissatisfaction caused by the opaque moderation explanations.

Additionally, opaque moderation occurred when YouTubers perceived differential treatments based on their fanbase. Some YouTubers claimed they experienced uneven punishments than other YouTubers, especially those largely subscribed by viewers. For example, one YouTuber replied to the other one who experienced demonetization due to inappropriate language use:

> Why can I find so many videos from these bigger channels that are violating guidelines? I can find an entire playlist of YouTuber A's videos that violate your

[15] https://support.google.com/youtube/answer/7083671?hl=en

> inappropriate language mark. It feels like these rules are being unevenly applied when small channels are mostly hit the bot.

Here, we used YouTuber A to refer to the YouTuber celebrity who was mentioned in this comment. In this quote, the YouTuber compared the thread poster's videos with several famous YouTubers' ones and described the observations where algorithmic punishments were imposed unequally between small and large video creators, also resonating with Caplan's findings [13]. The example here might involve personally uneven sentiment; however, this perceived unevenness also resonated with many media reports. For example, a large YouTuber, Logan Paul, had a problematic video that had remained on YouTube's recommendation trending before it was removed. Even this video was finally taken down by Logan himself [69]; his case showed the observed uneven moderation between YouTubers without sufficient explanations. These examples here reflected that limited transparency in algorithmic moderation caused it implausible for YouTubers to confirm or evaluate differential treatments but with confusion.

5.2 Managing Algorithmic Precarity

Algorithmic precarity refers to how algorithmic moderation engenders the work uncertainty of video content creation. Below is an example:

> My channel has 155,000 Subscribers, and over half of my videos were initially marked, lowering my ad revenue by close to 90%. I am sitting at 1/4th the revenue than before. Now I have gotten over 100 old videos approved, with one being marked not advertiser-friendly, as an example of how badly the bots are getting it wrong.

This YouTuber shared their experience of how algorithmic moderation led to existing financial loss, decreasing their normal revenue. While it was challenging to verify the exact numbers provided by the YouTuber, this quote did reveal the negative impacts of algorithmic moderation on profitability. This direct economic punishment could also indicate that algorithms mediate the work of video creation and cause uncertainty to work as a video content creator.

Video creation is time-consuming; YouTubers could spend 80 hours a week creating and editing videos [2]. Plus, video content is oftentimes time-sensitive. Thus, even temporary demonetization could incur a substantial financial loss even though they could reverse the demonetization ultimately. Two YouTubers wrote:

> My channel only gets 1 - 2 K views per video, and videos are all ad limited for the first 24 hours. So, by the time they manually approve my videos, I hardly get any more views.

> I have had my last 3-4 weeks' worth of uploads flagged and monetization restored on all of them after review. Unfortunately, that takes a few days, and I miss out on monetization for those days.

Both YouTubers shared similar experiences of losing progressive ad revenue because of waiting the time required to reverse algorithmic punishments manually. While their videos were turned back to monetize through the human review process, the waiting time of manual reviewing led to missing their deserved period for monetization. Thus, they failed to acquire the turn on their investments of time, energy, and ideas in video creation. This temporality feature of algorithmic moderation reflected the work precarity affecting video content creators' online life to real life.

Furthermore, algorithmic demonetization would further impose consequential impacts on YouTubers' real life. The punishments could lead YouTubers to experience various degrees of severity based on personal factors such as financial dependency on video creation or mental conditions. For example, one YouTuber shared their experience of account suspension along with the economic loss:

> I started this channel five years ago when my parents lost their jobs, so I could afford to go to college. It is hard to watch five years of hard work go down the drain and seek ways to pay for college. Maybe it is just a coincidence people all got terminated under similar circumstances. Regardless, I am upset that mine was wrongfully terminated in the first place, and because of it, I am losing income every day that my channel is gone.

In this quote, the YouTuber financially depended on creating videos to earn daily revenue. However, after experiencing account termination, they encountered an unstable living status to afford the daily expenditure, worsening their socioeconomic conditions. This case showed how YouTube demonetized a YouTuber at a channel level. Even though we cannot verify this YouTuber's account information, the causal logic contained in this example resonated well with many other YouTubers' stated experiences. Hence, YouTube moderation's algorithmic penalties and the corresponding socioeconomic impacts rendered YouTubers' work precarity.

YouTubers also experienced mental stress because of demonetization in algorithmic moderation. One YouTuber wrote:

> At this point, I've lost over $ 10,000 in revenue. I've lost my motivation. I'm depressed and anxious as hell, and I'm sick and tired of waiting for this to get sorted. I've been a Youtuber since 2008 with various channels, and I've never felt this frustrating about YouTube just automatically demonetizing my videos.

By stating a precise amount of revenue deduction, this YouTuber encountered anxiety and felt less motivated for future video creation, which implicitly indicated their negative emotions toward YouTube moderation.

Punished YouTubers often seek support from the community. For example, YouTuber B and C commented on a thread where the YouTuber experienced account suspension:

> YouTuber B: I can understand your feelings. After all that hard work, losing something so precious can be difficult. But do fight it out and try to get your channel back. We are all the way with you.
>
> YouTuber C: I can feel your pain; my channel also got terminated for no reason. It had 46k subscribers, I have tweeted them, and they said they had passed my request to the policy team. I suggest you keep trying whatever you are doing to get your channel back. Good luck.

Both YouTubers showed empathy for the punished YouTubers and encouraged them to reverse the issued punishments consistently. YouTuber C provided personal experiences of how to cope with moderation penalties. Also, they both expressed emotional support and recommended the punished YouTuber to consistently reverse algorithmic penalties. This instance reflected that both YouTubers had a high tolerance towards algorithmic moderation. With feeling a lack of agency, they had hope for a better moderation system.

However, experiencing all these aspects of algorithmic precarity, YouTubers also showed how they were empowered by such precarity to exert the autonomy of stabilizing their income. They diversified their income not solely to rely on ad income. Specifically, we found that they

collaborated with various funding platforms or marketing sponsors [105] to acquire direct funding, alleviating future demonetization impacts. One YouTuber suggested:

> Try to seek monetization directly from your audience (e.g., Patreon, Ko-fi, etc.) Honestly, I can't recommend any creator pin all their revenue on AdSense, even if they are not at risk of being removed from the YPP.

In this quote, YPP refers to the YouTuber Partner Program. The YouTuber mentioned 'Patreon' and 'Ko-fi'; both are external crowdfunding platforms for content creators like YouTubers to establish membership with audiences by posting exclusive content to acquire direct monetary donations. Here, the discussions and behaviors about using external funding platforms indicated that YouTubers actively coped with the future financial loss to both exert their autonomy and alleviate their work uncertainty.

Some YouTubers described how they could reverse the demonetization penalty by contacting Multi-Channel Networks (MCNs), as one YouTuber wrote:

> MCNs can sometimes have contact with YouTube to assist with enabling monetization on channels that might normally be refused.

Here, MCNs refer to organization partners who help establish brands for YouTubers and marketing their videos [37]. Contacting outside entities reflected the YouTuber actively coped with penalties and pointed out the generally limited communicative methods for YouTubers to solve their economic problems with the YouTube platform.

Besides YouTubers showing how they coped with the demonetization and its socioeconomic impacts, they also switched between or migrated to other streaming platforms to seek a lower degree of future moderation risks, different from prior work uncovering that users stayed at the same platform to do so [16]. For example, one YouTuber who experienced algorithmic punishments wrote in the thread:

> In any case, I am going to try Bitchute for now. It has support for people who are streamers. For me, what is happening with YouTube right now is the final straw, and I'm tired of that platform and want to go somewhere else.

This YouTuber mentioned their desire for completely transferring to an alternative video streaming platform of YouTube, namely Bitchute, to seek lower moderation. They described the moderation experience by claiming how YouTube rarely showed considerations for YouTubers. However, here we need an analytical lens to see such platform shift actions. Since YouTube is currently the largest video streaming platform, competitors hardly have more userbase and daily engagement data than YouTube. Hence, YouTubers shifting between platforms would encounter challenges for acquiring more future economic benefits due to the lower fanbase, incomparable to what they can reach on YouTube.

5.3 Learning and Applying Algorithmic Know-How

YouTubers collectively made sense of algorithmic punishments, developing and disseminating algorithmic know-how or practical knowledge of YouTube moderation. Specifically, YouTubers shared and analyzed their punishment experiences to speculate about moderation algorithms, which in turn informed operations of repairing their past moderation punishments and avoiding future interactions with YouTube moderation. For example, two YouTubers discussed in their dialogue:

> YouTuber D: The bot scanned my pre-stream has concluded that it is not suitable for ads. I have no strikes on my account, and I have had no unsuccessful reviews. What is it basing these scans on? How is it predicting the future, and so badly?
>
> YouTuber E: Your tags, title, description, and thumbnail are all available pre-stream. Maybe that's what the bot guessed from.

In the quotes, 'pre-stream' refers to the published live streaming on YouTube Live that is neither scheduled nor held yet. YouTuber D expressed the confusion of how YouTube's algorithms made moderation decisions without reviewing the stream content. YouTube E stated that metadata of videos helped the bot make algorithmic predictions/classifications. Also, YouTuber E used the words 'guess' and 'maybe' to describe the bot's classifying mechanism, expressing uncertainty in their analysis of moderation. This dialogue resonated with Caplan's findings [13] that the video's metadata might trigger YouTube moderation. It further showed the process of collective sensemaking: how YouTubers collaborated to piece their past experiences together and speculated about how moderation algorithms worked.

YouTubers applied the know-how about metadata moderation in their content creation. Many YouTubers discussed how they could self-moderate their metadata to repair punishments. One YouTuber wrote:

> I've tweaked my video title and thumbnail over the course of a week, and I've had my status automatically changed back to monetization.

The YouTuber mentioned that by editing metadata (e.g., tags, descriptions, thumbnails, user comments, and titles), they changed their past videos to repair demonetization penalties. Resonating with self-moderation on social media [85], this example indicated a practical level of self-moderation behaviors on YouTube. Also, both examples regarding tweaking metadata further pointed out the labor of dealing with moderation punishments.

Even holding such algorithmic knowledge, some YouTubers felt powerless to avoid future moderation risks when massive audience comments trigger demonetization punishments. One YouTuber mentioned:

> The problem is that advertisers didn't like that those [hateful] comments were there. The problem is not that YouTube doesn't deal with them; it's that YouTube cannot police every comment posted. What they need is a quality filter, just like Twitter. Accounts with verified phone numbers, emails, maybe throw in a Facebook link to secure it that little bit further.

In this example, the YouTuber envisioned a filter to manage the audience's comments on their videos. Three layers of information were disclosed here. First, resonating with prior studies regarding bad actors gaming algorithmic systems [23,47] and the third YouTube Adpocalypse's origins, YouTubers try to mitigate moderation risks brought by problematic audience behaviors. Second, even though YouTube had provided keywords blocking functions for YouTubers to filter problematic audience comments, the example here showed that it still provided insufficient support to moderate potential commentators. Last, the YouTuber here called for verified audiences under the assumption that non-qualified audiences can still watch videos for their monetization. This example further showed that YouTubers having algorithmic knowledge could provide reflective suggestions for moderation systems' design even though they are in the hardship of avoiding future moderation risks.

Furthermore, part of the algorithmic know-how was to strategically use the appeal process when YouTubers deemed the algorithmic decisions inaccurate. For example, one YouTuber who experienced account suspension shared their strategy:

> It's been about two months since this took place, and I figured I should update you on how things are going. So, I eventually got my channel back up, I contacted Trusted Flagger on Twitter, and he helped me get the strikes taken off my account, and it was back up and running by the following week.

This YouTuber provided an alternative way of the appeal process and remedying past punishments: using an external social media platform to contact the YouTube staff. On the one hand, this action overcame the limitation of an appeal process that YouTube allows YouTubers to do only once. On the other, this information implicitly reflected a lack of a communicative path for YouTubers to contact the YouTube platform for moderation issues directly on YouTube.

Knowing that appeal, even if effective, would take time. YouTubers thus adapted their work schedules to cope with such time costs. One YouTuber wrote:

> Most appeals are handled within a day. And that appeal can be handled before the content goes live, so I Just offset my schedule by 2-3 days. If I upload daily, I upload my Wednesday content on Monday, schedule, and appeal if it is hit.

In this quote, the YouTuber described their action of avoiding potential financial loss by posting live streaming schedules earlier. Behind this process, YouTubers utilized the time that an appeal usually took to have enough time to prepare for future moderation.

Reflexively, YouTubers pointed to how their feedback in the form of appeal could help improve moderation algorithms. Below is a conversation between two YouTubers:

> YouTuber F: I've had one video in my last 30 uploads not be ad limited. The rest I have had to send away for review. FYI all live streams where I get ad limited on in the past have been successfully reviewed afterward.
>
> YouTuber G: That's good! You are helping train the bot by doing this. If the bot isn't able to make a decision because there is too little for it to go off of, or it doesn't know what to do with some of the data, it hits the video to be safe.

YouTuber F, in detail, described their demonetization experiences as well as the appeal's outcomes. YouTuber G explicitly explained the reason why the bot makes demonetization decisions and highlighted the expectation that more appeals can help the algorithms be trained better, which was encouraged by YouTuber G to repair past penalties. Besides, they further developed the prior knowledge from the bot algorithmically classifying content based on metadata to how the bot made final decisions for issuing penalties. This dialogue here showed both YouTubers' desire for a moderation system with better perceived accuracy.

6 DISCUSSION

This study analyzed how YouTubers interact with the algorithmic moderation on YouTube, as summarized in Table 1. We extended prior scholarship by uncovering moderation systems' socioeconomic penalties and effects and how YouTubers collectively perceived, learned from, and coped with these algorithmic punishments. In this section, we will discuss YouTubers' algorithmic precarity and their labor of repairing and avoiding algorithmic moderation. We will show how post-punishment peer and platform support could serve as a restorative justice means. Ultimately, we call for trustful explanations of algorithmic moderation, compensating

economic loss for falsely demonetized YouTubers, and algorithmic moderation systems with more transparency.

Table 1. Socioeconomic Interactions with YouTube's Algorithmic Moderation

Categories	Themes
Opaque algorithmic moderation failed to support user's reasoning	• YouTubers did not receive notifications for specific moderation punishments • Normal audiences with few activities got undeserved moderation punishments without convincing or sufficient explanations. • The unfair moderation was perceived to impose on YouTubers disproportionally compared with the larger YouTubers. • The moderation explanations also cannot provide enough credibility to YouTubers (e.g., what video violated community guidelines in specific timestamps). • YouTubers personally experienced inconsistent or irregular moderation without enough explanations.
The algorithmic precarity as YouTubers	• Future ad income decreased or is deprived (i.e., demonetization) due to YouTube moderation. • YouTubers exerted their autonomy to stabilize the income by associating with multiple monetization/crowdfunding platforms. • YouTubers switched between different streaming platforms to avoid exclusively leaning on YouTube. • YouTubers used different communication mediums to contact YouTube Team due to such limited provision.
Learning and applying algorithmic knowledge to repair and avoid moderation	• YouTubers collectively theorized, shared, and practiced knowledge that various types of metadata trigger YouTube algorithmic moderation. • YouTubers collectively shared and practiced how they can and cannot handle past moderation punishments and avoid future moderation. • YouTubers collectively theorized how YouTube algorithms make moderation decisions.

6.1 The Labor of Dealing with Moderation Algorithms and Socioeconomic Punishments

Previous scholarship has examined the labor of human moderators in managing and curating content and enforcing norms [28,101]. Our findings centered on the other side of moderation and shed light on the labor of video content creators, the moderated, in coping with their punishments. We showed how moderation algorithms intersect with YouTubers' content creation work, engendering a necessary form of algorithmic labor to comply with moderation algorithms on YouTube and make their videos "advertiser-friendly."

Specifically, moderation algorithms have conditioned YouTubers' content creation labor in several ways. YouTube's algorithmic moderation has resulted in more work from YouTubers. As YouTube moderation develops complex policies and enforcement strategies [13], its opacity grows. Subsequently, YouTubers must do more work to restore transparency to their punishments. Prior work has discussed how social media users typically encountered algorithmic moderation's inexplicability (e.g., [33,47,52,74]). In the case of YouTube moderation, the opacity has multiple layers. We showed that there was a lack of warning prior to YouTube issuing algorithmic penalties. When punishments happened, their associated explanations seemed to be machine-generated with generic and insufficient language. Previous research has confirmed the importance of human moderators in providing moderation explanations [47,101]. In our study, YouTubers who were unsatisfied with automatic explanations had to do the work of appeal. There was also a certain degree of inexplicability in executing algorithmic decisions between small and large YouTubers. Lastly, the opacity also manifested in the lack of direct

communicative methods to apply for human interventions on issued algorithmic punishments, unlike other social media that allowed users to directly contact moderators [39,52]. Each layer of opacity would require a sizable effort for YouTubers to wrestle with.

As cultural production moves onto platforms, communication scholars have observed how algorithms exacerbate the instability of such work [29]. In particular, our study showed how the opacity of moderation algorithms had intensified the precarity of YouTubers' content creation work. Besides striving to create popular content, they also have to be mindful of the inconsistency of moderation algorithms, where punishments were oftentimes issued irregularly on the same video content or disproportionally on different YouTubers due to their scale of the fan base. Such inconsistency resonated with Vaccaro et al.'s investigations on Facebook, where policies were found to be applied on users at uneven levels [95]. Also, our study extended Gillespie's work that unevenness not only appeared in distributed human moderation [39] but also in automated moderation, as our case of YouTube's demonetization penalties shown.

Thus, part of YouTubers' work is to manage such algorithmic precarity. Previous scholarship has reported how YouTubers have to consider financial alternatives to ad revenue from YouTube [4,13]. More broadly, managing algorithmic precarity. Besides the financial strategy, our findings pointed to several other dimensions of managing algorithmic precarity. First, YouTubers have to both repair their past moderation punishments and avoid future moderation risks. Extending prior work discussing how users circumvented moderation on the original social media platforms [16,38], we found YouTubers become skilled at editing content such as metadata of videos to exert self-moderation. They use different scheduling to work with the delay of demonetization (future moderation) as well as consistently and strategically appeal to reinstate monetization status (past moderation).

Second, YouTubers engage in social practices to manage algorithmic precarity. While post-punishment transparency work denotes short-term effort to repair, YouTubers turn to online communities like our study site for various forms of support. They have to engage in long-term learning, especially when they believe that moderation algorithms are also learning and evolving. They utilize online communities to constantly update their algorithmic know-how as a preemptive strategy, as well as remain reflexive. Some even organized collective actions to unionize and challenge the platform's governance decisions [68]. In this regard, our study provided a detailed account of how moderation algorithms are also intertwined with the already fraught labor relation of content creation.

Furthermore, there is an affective dimension to precarity management. People could experience anxiety, loss of agency, and negative emotions when dealing with complex algorithmic systems [12,49]. YouTubers would experience intense negative emotions upon demonetization punishments because these decisions are of high stakes and also because these decisions bring high uncertainty and opacity.

Lastly, the work precarity resulting from demonetization punishments also has transferable meanings. On the one hand, prior studies largely investigated how users who mainly generate text content interact with moderation systems on social media such as Reddit [48,52], Instagram [16,30], and Facebook [66,93]. We extended these studies by focusing on video content creators' perspectives and stressed the increased precarity due to socioeconomic punishments: demonetization, its impacts, and the labor of handling moderation punishments. On the other, this study indirectly extended the concept of precariousness mainly discussed around sharing economy and digital worksites (e.g., [60,64,76,79]). We can see that content creation involves monetization and its connections with audiences (e.g., crowdfunding from subscribers on

Twitch [51,71]). At the same time, we should note the precarity caused by socioeconomic content moderation and how creators manage such precarity, just as the case of YouTubers shown.

As such, YouTubers' algorithmic labor involves striking a delicate balance between enhancing visibility on the one hand and avoiding moderation algorithms on the other. Due to this, the socioeconomic punishment on YouTube could be considered as qualitatively more severe than free expression platforms such as Twitter and Reddit: punishment comes with not just the deprivation of the privilege of expression but also the deprivation of revenue.

6.2 Post-Punishment Support as Restorative Justice

In HCI and CSCW, researchers have reflected upon the limits of retributive justice, the punitive system that YouTube currently employs, and paid more attention to restorative justice [6,81]. Restorative justice denotes "the repair of justice through reaffirming a shared value-consensus in a bilateral process" [99] and values processes of healing and reconciliation. A restorative justice lens implies that we should also value offenders' experiences and seek to repair harm and reintegrate offenders back into the community [6]. Relatedly, post-punishment support could serve as a restorative justice means. Specifically, our findings highlighted two existing forms of support: peer support and platform support.

Prior studies discussed peer support on social media [58,61,102]. In the context of YouTube, we found that YouTubers shared and analyzed their experiences or knowledge to speculate about moderation algorithms. Previous research discussed that social media users hardly felt a sense of agency in algorithmic moderation systems [32] and how they took actions to strengthen their agency [66]. By extending these studies, we found YouTubers actively learned from algorithmic penalties and developed knowledge of moderation algorithms collectively. When platform policies are unclear and enforced unevenly [13], algorithmic know-how becomes important situated knowledge in algorithmically mediated work.

Importantly, we uncovered that peer support is not only for circumventing future moderation. Prior studies have investigated how punished users supported each other to bypass future moderation on social media [16,38]. However, YouTubers, as content creators, were found to both repair past moderation punishments and avoid future moderation risks. Sometimes, even though being aware of practical knowledge of moderation algorithms, they might fail to circumvent specific future punishments. Due to algorithmic precarity and its affective ramifications, emotional support is another category of social support we observed in the community. YouTubers jointly expressed emotional support for those who experienced penalties in their precarious work of content creation and sharing.

Besides peer support, YouTubers also value support from the platform. YouTubers highlighted the importance of direct communication with the platform. They turned to external communication platforms seeking informational support and proactively bridged the communication with YouTube. However, they are not content with the current level of platform support. Previous research uncovered that users could hardly communicate with the platform for moderation issues [46] and called for a more communicative process in moderation systems [93]. YouTube was found to have similar problems. YouTubers in our study were unsatisfied with their communication with the platform and voiced complaints about the lack of transparency even within the appeal process.

Both peer support and platform support represent a meaningful departure from retributive justice's simplistic logic of using penalties to fix offenders. When offenders do not comprehend

the penal rationales, punishments alone could hardly reform offenders. West proposed that content moderation systems could be more educational [66]. We extend this by highlighting how peer and platform support could also contribute to offenders' behavioral improvement.

6.3 Design Considerations

Previous researchers have called for democratic accountability, transparency, and free expression from the internet platforms [39,48,66]. Reflecting upon the opacity of YouTube's algorithmic moderation, we suggest that YouTube could offer convincing or sufficient explanations at the user level. For example, suppose a YouTuber is demonetized by the thumbnail of their videos. In that case, the algorithmic system should indicate the punishment decision is from the thumbnail rather than similarly stating which policy is violated. Resonating with prior evidence that users might rush to make conclusions based on personal experiences [87], we stress the importance of providing educational explanations to match with various YouTubers' backgrounds by the algorithmic moderation system. Furthermore, given the findings that YouTubers felt unsatisfied with the appeal procedure, we argue that YouTube could provide sufficient social support to human moderatos to reach considered moderation decisions [78,93], which could potentially improve the transparency of moderation.

Moreover, YouTube could learn more from YouTubers for the better accuracy of algorithmic moderation. Our findings involving content policies, human moderators, and YouTubers manifested YouTube's algorithmic decision-making processes have remained confusing to YouTubers. Hence, we called for YouTube's algorithms to learn from YouTubers more comprehensively, where YouTube's algorithms/classifiers can include records of YouTubers' historical content as a factor to predict penalty decisions for future videos as a reference. We found many YouTubers with all past nonproblematic records of videos but suddenly encountered algorithmic penalties, rendering an inconsistent moderation procedure. Even though reversing those penalties ultimately, they took the economic loss that happened at that time. Involving this historical factor could help the classifiers receive more training data from YouTubers to reach more accurate and confident results. Supplementing this design, moderation systems should also be aware of videos that already pass moderation since we found YouTubers reported moderation systems repeatedly tagged them with 'limited ads.'

We call for attention to potential falsely moderated YouTubers' subjective experiences, drawing from a legal discussion of how a successful justification of penalty should consider policy offenders' subjective experiences [55]. As we described in Section 5.2, YouTubers sensed various degrees of severity from penalties given their financial dependency on video creation or mental conditions. Hence, the YouTube platform could consider compensating YouTubers for falsely executed YouTube moderation. For example, content policies could articulate that after a successful human review (appeal) process, YouTubers could acquire the ad income calculated by the time from issuing demonetization to successfully passing the appeal procedure. This change can potentially allow YouTube's algorithmic moderation to be more accountable for users and show their enterprise responsibility.

To sum up, this study of YouTube moderation could also provide transferable design implications for other social media. For platforms providing video monetization from ads/advertising such as Facebook [31] and Twitter [94], moderation explanations could provide more actionable suggestions such as editing/repairing specific timestamps of videos to comply with community guidelines instead of largely leaning on the human review (appeal) to solve false-negative algorithmic moderation decisions. This could make algorithmic decisions more

explainable. Also, platforms such as Facebook, Twitter, and YouTube benefit from business interests of providing audience management functions such as 'customize audience' to gain more income. However, issuing moderation punishments on content creators, platforms have scarcely considered the possibility of audiences' problematic behaviors, as Section 5.3 shown. Thus, we also called for the potential functionality of audiences' identity qualification to improve true-positive portions of algorithmic moderation decisions by involving outliers in the training process, referring to problematic viewer behaviors.

7 LIMITATIONS AND FUTURE WORK

Researchers in the CSCW community have previously used specific subreddits' discussion data for different research topics (e.g., [5,80]). We acknowledge using this type of data could be possible to bring bias, where YouTubers might share biased or conflicting information. However, we can still systematically distill meaningful experiential conclusions from the data and point out the urgent importance of designing more transparent algorithmic moderation to mitigate such bias. Even an ideal machine learning model cannot reach 100% test accuracy due to various reasons (e.g., overfitting) [24]. So, even though YouTubers' content is classified as true-positive problematic, when YouTube cannot provide convincing explanations, YouTubers cannot learn from their past behaviors. Thus, the bias will last. Also, it is possible that people who find moderation punishments to be fair don't come to the subreddit to complain, but this doesn't make the subreddit data less valuable. Rather, it provides a window in which frustrations lie and from where design could bring consistent punishment experiences not to be frustrating. So, given the big data nature of YouTube videos[16], we aim to understand YouTubers' lived experience of algorithms in content moderation. We treat this study as an exploratory study of YouTube moderation and present balanced interpretations from the Reddit data.

This research only collected the data that explicitly discussed YouTube's content moderation based on relevant literature and media reports. Hence, those implicit expressions that do not contain relevant keywords are hard to identify. Also, the dataset does not include information about YouTuber categories (e.g., games, lifestyle, music). As such, in the analysis process, we cannot unearth whether differences of experiences or post-punishment interactions with moderation would exist between specific YouTuber types. Besides, we did not discuss human moderation (i.e., initially manual flag or moderation) on YouTube due to YouTube's high (>95%) dependence on automatic moderation [42].

Surveys or interviews with YouTubers would reveal YouTubers' in-depth understandings and considerations of moderation on YouTube. There could be future studies focusing on how YouTubers in specific categories experience moderation through methods such as interviews, surveys, or analyzing YouTube video data. This would potentially in-depth depict how YouTubers theorize the mechanism of YouTube's moderation systems and how they perceive the systems' fairness and sufficiency of moderation explanations.

8 CONCLUSION

As online platforms like YouTube carry growing significance in people's socioeconomic life, moderation plays a profound role in dictating some's livelihoods. What this study showed is how the opacity of algorithmic moderation, existing at multiple layers, injects more precarity in

[16] https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute/

YouTubers' labor. They are in need of peer and platform support that could come in a sufficient and efficient way. The restorative perspective could help yield meaningful insights into how post-punishment support could be envisioned and designed into existing moderation systems.

ACKNOWLEDGMENTS

We thank the associate chairs and anonymous reviewers for their insightful feedback and suggestions. This work is partially supported by the NSF, under grant no. 2006854.

REFERENCES

[1] Julia Alexander. 2018. What is YouTube demonetization? An ongoing, comprehensive history. Polygon. Retrieved from https://www.polygon.com/2018/5/10/17268102/youtube-demonetization-pewdiepie-logan-paul-casey-neistat-philip-defranco

[2] Julia Alexander. 2019. The golden age of YouTube is over. The Verge. Retrieved from https://www.theverge.com/2019/4/5/18287318/youtube-logan-paul-pewdiepie-demonetization-adpocalypse-premium-influencers-creators

[3] Julia Alexander. 2019. YouTube says it has 'no obligation' to host anyone's video. The Verge. Retrieved from https://www.theverge.com/2019/11/11/20955864/youtube-terms-of-service-update-terminations-children-content-ftc

[4] Ali Alkhatib and Michael Bernstein. 2019. Street–level algorithms: A theory at the gaps between policy and decisions. In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), Association for Computing Machinery, New York, NY, USA, 1–13. DOI:https://doi.org/10.1145/3290605.3300760

[5] Maria Antoniak, David Mimno, and Karen Levy. 2019. Narrative paths and negotiation of power in birth stories. Proceedings of the ACM on Human-Computer Interaction 3, 1–27. DOI:https://doi.org/10.1145/3359190

[6] Mariam Asad. 2019. Prefigurative design as a method for research justice. Proc. ACM Human-Computer Interact. 3, CSCW (November 2019), 41. DOI:https://doi.org/10.1145/3359302

[7] Joan Isaac Biel and Daniel Gatica-Perez. 2011. VlogSense: Conversational behavior and social attention in YouTube. ACM Transactions on Multimedia Computing, Communications and Applications 7 S, 1–21. DOI:https://doi.org/10.1145/2037676.2037690

[8] Sophie Bishop. 2019. Managing visibility on YouTube through algorithmic gossip. New Media Soc. 21, 11–12 (November 2019), 2589–2606. DOI:https://doi.org/10.1177/1461444819854731

[9] Sophie Bishop. 2020. Algorithmic Experts: Selling Algorithmic Lore on YouTube. Soc. Media + Soc. 6, 1 (2020), 205630511989732. DOI:https://doi.org/10.1177/2056305119897323

[10] Hannah Bloch-Wehba. 2020. Automation in Moderation. Cornell Int. Law J. 53, (2020), 41.

[11] Glenn A. Bowen. 2008. Naturalistic inquiry and the saturation concept: a research note. Qual. Res. 8, 1 (February 2008), 137–152. DOI:https://doi.org/10.1177/1468794107085301

[12] Taina Bucher. 2017. The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. Inf. Commun. Soc. 20, 1 (January 2017), 30–44. DOI:https://doi.org/10.1080/1369118X.2016.1154086

[13] Robyn Caplan and Tarleton Gillespie. 2020. Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy. Soc. Media + Soc. 6, 2 (2020). DOI:https://doi.org/10.1177/2056305120936636

[14] Mark Carman, Mark Koerber, Jiuyong Li, Kim Kwang Raymond Choo, and Helen Ashman. 2018. Manipulating Visibility of Political and Apolitical Threads on Reddit via Score Boosting. In Proceedings - 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering, Trustcom/BigDataSE 2018, Institute of Electrical and Electronics Engineers Inc., 184–190. DOI:https://doi.org/10.1109/TrustCom/BigDataSE.2018.00037

[15] Bobo Chan. 2020. Is Being A YouTuber Still Lucrative? Jumpstart Magazine. Retrieved from https://www.jumpstartmag.com/is-being-a-youtuber-still-lucrative/

[16] Stevie Chancellor, Jessica Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. #thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16, ACM Press, New York, New York, USA. DOI:https://doi.org/http://dx.doi.org/10.1145/2818048.2819963

[17] Minnie Che. 2019. The New Job Market is Online: YouTube. OnLabor. Retrieved from https://onlabor.org/the-new-job-market-is-online-youtube/

[18] Jennifer Cobbe. 2020. Algorithmic Censorship by Social Platforms: Power and Resistance. Philos. Technol. (October 2020), 1–28. DOI:https://doi.org/10.1007/s13347-020-00429-0

[19] Kelley Cotter. 2019. Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. New Media Soc. 21, 4 (April 2019), 895–913. DOI:https://doi.org/10.1177/1461444818815684

[20] Kate Crawford and Tarleton Gillespie. 2016. What is a flag for? Social media reporting tools and the vocabulary of complaint. New Media Soc. 18, 3 (March 2016), 410–428. DOI:https://doi.org/10.1177/1461444814543163
[21] John B. Davis and Wilfred Dolfsma. 2008. The Elgar Companion to Social Economics.
[22] Caitlin Dewey. 2016. Why YouTubers are accusing the site of rampant 'censorship.' The Washington Post. Retrieved from https://www.washingtonpost.com/news/the-intersect/wp/2016/09/01/why-youtubers-are-accusing-the-site-of-rampant-censorship/
[23] Nicholas Diakopoulos. 2016. Accountability in algorithmic decision making. Commun. ACM 59, 2 (February 2016), 56–62. DOI:https://doi.org/10.1145/2844110
[24] Pedro Domingos. 2012. review articles Tapping into the "folk knowledge" needed to advance machine learning applications. Commun. ACM 55, 10 (2012). DOI:https://doi.org/10.1145/2347736.2347755
[25] Nicola Döring and M. Rohangis Mohseni. 2020. Gendered hate speech in YouTube and YouNow comments: Results of two content analyses. Stud. Commun. Media 9, 1 (March 2020), 62–88. DOI:https://doi.org/10.5771/2192-4007-2020-1-62
[26] Nicola Döring and M Rohangis Mohseni. 2019. Communication Research Reports Fail videos and related video comments on YouTube: a case of sexualization of women and gendered hate speech? Commun. Res. Reports 36, 3 (2019), 254–264. DOI:https://doi.org/10.1080/08824096.2019.1634533
[27] Finale Doshi-Velez and Been Kim. 2017. Towards A Rigorous Science of Interpretable Machine Learning. arXiv Prepr. (2017). Retrieved from https://arxiv.org/abs/1702.08608v2
[28] Bryan Dosono and Bryan Semaan. 2019. Moderation practices as emotional labor in sustaining online communities: The case of AAPI identity work on reddit. CHI Conf. Hum. Factors Comput. Syst. Proc. (CHI 2019) (May 2019), 1–13. DOI:https://doi.org/10.1145/3290605.3300372
[29] Brooke Erin Duffy. 2020. Algorithmic precarity in cultural work. Commun. Public 5, 3–4 (September 2020), 103–107. DOI:https://doi.org/10.1177/2057047320959855
[30] Stefanie Duguay, Jean Burgess, and Nicolas Suzor. 2020. Queer women's experiences of patchwork platform governance on Tinder, Instagram, and Vine. Converg. Int. J. Res. into New Media Technol. 26, 2 (April 2020), 237–252. DOI:https://doi.org/10.1177/1354856518781530
[31] Maria Dykstra. 2019. How to Monetize Your Facebook Video With Facebook Ad Breaks. Social Media Examiner. Retrieved from https://www.socialmediaexaminer.com/how-to-monetize-facebook-video-facebook-ad-breaks/
[32] 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, Association for Computing Machinery, New York, NY, USA, 2371–2382. DOI:https://doi.org/10.1145/2858036.2858494
[33] Motahhare Eslami, Kristen Vaccaro, Karrie Karahalios, and Kevin Hamilton. 2017. "Be careful; things can be worse than they appear": Understanding Biased Algorithms and Users' Behavior around Them in Rating Platforms. In ICWSM, 62–71. Retrieved from www.aaai.org
[34] Megan Farokhmanesh. 2018. YouTube is demonetizing some LGBT videos — and adding anti-LGBT ads to others. The Verge. Retrieved from https://www.theverge.com/2018/6/4/17424472/youtube-lgbt-demonetization-ads-algorithm
[35] Joan Feigenbaum, Aaron D. Jaggard, and Rebecca N. Wright. 2011. Towards a formal model of accountability. In Proceedings New Security Paradigms Workshop, ACM Press, New York, New York, USA, 45–55. DOI:https://doi.org/10.1145/2073276.2073282
[36] Casey Fiesler, Jialun Aaron Jiang, Joshua McCann, Kyle Frye, and Jed R. Brubaker. 2018. Reddit rules! Characterizing an ecosystem of governance. 12th Int. AAAI Conf. Web Soc. Media, ICWSM 2018 (2018), 72–81.
[37] Jacob Gardner and Kevin Lehnert. 2016. What's new about new media? How multi-channel networks work with content creators. Bus. Horiz. 59, 3 (May 2016), 293–302. DOI:https://doi.org/10.1016/j.bushor.2016.01.009
[38] Ysabel Gerrard. 2018. Beyond the hashtag: Circumventing content moderation on social media. New Media Soc. 20, 12 (December 2018), 4492–4511. DOI:https://doi.org/10.1177/1461444818776611
[39] Tarleton Gillespie. 2018. Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press. Retrieved from https://www.degruyter.com/document/doi/10.12987/9780300235029/html
[40] Olivia Goldhill. 2020. How YouTube shields advertisers (not viewers) from harmful videos. Retrieved from https://qz.com/1785613/how-youtube-shields-advertisers-not-viewers-from-harmful-videos/
[41] Shraddha Goled. 2020. Online Content Moderation: To AI or Not. Analytics India Magazine. Retrieved January 15, 2021 from https://analyticsindiamag.com/online-content-moderation-to-ai-or-not/
[42] Google. 2021. YouTube Community Guidelines enforcement. Retrieved March 16, 2021 from https://transparencyreport.google.com/youtube-policy/removals?hl=en
[43] Robert Gorwa, Reuben Binns, and Christian Katzenbach. 2020. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. Big Data Soc. 7, 1 (January 2020), 205395171989794.

DOI:https://doi.org/10.1177/2053951719897945

[44] James Grimmelmann. 2015. The Virtues of Moderation. Yale J. Law Technol. 17, (2015). Retrieved from https://heinonline.org/HOL/Page?handle=hein.journals/yjolt17&id=42&div=&collection=

[45] Simon Niklas Hellmich. 2017. What is socioeconomics? An overview of theories, methods, and themes in the field. Forum Soc. Econ. 46, 1 (2017), 3–25. DOI:https://doi.org/10.1080/07360932.2014.999696

[46] 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. Proc. ACM Human-Computer Interact. 3, CSCW (November 2019), 1–33. DOI:https://doi.org/10.1145/3359294

[47] Shagun Jhaver, Iris Birman, Eric Gilbert, and Amy Bruckman. 2019. Human-machine collaboration for content regulation: The case of reddit automoderator. ACM Trans. Comput. Interact. 26, 5 (July 2019), 1–35. DOI:https://doi.org/10.1145/3338243

[48] Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2019. Does transparency in moderation really matter?: User behavior after content removal explanations on reddit. Proc. ACM Human-Computer Interact. 3, CSCW (2019). DOI:https://doi.org/10.1145/3359252

[49] Shagun Jhaver, Yoni Karpfen, and Judd Antin. 2018. Algorithmic anxiety and coping strategies of airbnb hosts. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, New York, USA, 1–12. DOI:https://doi.org/10.1145/3173574.3173995

[50] Lin Jin. 2020. The Creator Economy Needs a Middle Class. Harvard Business Review. Retrieved from https://hbr.org/2020/12/the-creator-economy-needs-a-middle-class

[51] Mark R. Johnson and Jamie Woodcock. 2019. "And Today's Top Donator is": How Live Streamers on Twitch.tv Monetize and Gamify Their Broadcasts. Soc. Media + Soc. 5, 4 (October 2019), 205630511988169. DOI:https://doi.org/10.1177/2056305119881694

[52] Prerna Juneja, Deepika Rama Subramanian, and Tanushree Mitra. 2020. Through the looking glass: Study of transparency in Reddit's moderation practices. Proc. ACM Human-Computer Interact. 4, GROUP (January 2020), 1–35. DOI:https://doi.org/10.1145/3375197

[53] Rene F. Kizilcec. 2016. How much information? Effects of transparency on trust in an algorithmic interface. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 2390–2395. DOI:https://doi.org/10.1145/2858036.2858402

[54] Brian Koerber. 2018. Here's why Logan Paul's video showing suicide is so dangerous. Mashable. Retrieved from https://mashable.com/2018/01/02/logan-paul-suicide-video-explainer/

[55] Adam J. Kolber. 2009. The Subjective Experience of Punishment. Columbia Law Rev. 109, (2009). Retrieved from https://heinonline.org/HOL/Page?handle=hein.journals/clr109&id=186&div=&collection=

[56] Susanne Kopf. 2020. "Rewarding Good Creators": Corporate Social Media Discourse on Monetization Schemes for Content Creators. Soc. Media + Soc. 6, 4 (October 2020), 205630512096987. DOI:https://doi.org/10.1177/2056305120969877

[57] Yubo Kou and Xinning Gui. 2020. Mediating Community-AI Interaction through Situated Explanation. Proc. ACM Human-Computer Interact. 4, CSCW2 (October 2020), 1–27. DOI:https://doi.org/10.1145/3415173

[58] Cliff Lampe, Jessica Vitak, Rebecca Gray, and Nicole B. Ellison. 2012. Perceptions of Facebook's value as an information source. In Proceedings of the SIGCHI conference on human factors in computing systems, ACM Press, New York, New York, USA, 3195–3204. DOI:https://doi.org/10.1145/2207676.2208739

[59] Ralph LaRossa. 2005. Grounded Theory Methods and Qualitative Family Research. J. Marriage Fam. 67, 4 (November 2005), 837–857. DOI:https://doi.org/10.1111/j.1741-3737.2005.00179.x

[60] Min Kyung Lee, Daniel Kusbit, Evan Metsky, and Laura Dabbish. 2015. Working with machines: The impact of algorithmic and data-driven management on human workers. In Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 1603–1612. DOI:https://doi.org/10.1145/2702123.2702548

[61] Yong Liu, Jayant Venkatanathan, Jorge Goncalves, Evangelos Karapanos, and Vassilis Kostakos. 2014. Modeling what friendship patterns on facebook reveal about personality and social capital. ACM Trans. Comput. Interact. 21, 3 (June 2014), 1–20. DOI:https://doi.org/10.1145/2617572

[62] Emma J Llansó. 2020. No amount of "AI" in content moderation will solve filtering's prior-restraint problem. Big Data Soc. 7, 1 (January 2020), 205395172092068. DOI:https://doi.org/10.1177/2053951720920686

[63] Arwa Mahdawi. 2017. PewDiePie thinks "Death to all Jews" is a joke. Are you laughing yet? The Guardian. Retrieved from https://www.theguardian.com/commentisfree/2017/feb/15/youtube-pewdiepie-thinks-death-to-all-jews-joke-laughing-yet

[64] D. Michael O'Regan and Jaeyeon Choe. 2017. Airbnb: Turning the Collaborative Economy into a Collaborative Society. . Springer, Cham, 153–168. DOI:https://doi.org/10.1007/978-3-319-51799-5_9

[65] Melissa J. Morgans. 2017. Freedom of Speech, the War on Terror, and What's YouTube Got to Do with It: American Censorship during Times of Military Conflict. Fed. Commun. Law J. 69, (2017). Retrieved from

https://heinonline.org/HOL/Page?handle=hein.journals/fedcom69&id=163&div=&collection=

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

[67] Casey Newton. 2019. The Terror Queue: These moderators help keep Google and YouTube free of violent extremism — and now some of them have PTSD. The Verge. Retrieved from https://www.theverge.com/2019/12/16/21021005/google-youtube-moderators-ptsd-accenture-violent-disturbing-content-interviews-video

[68] Valentin Niebler. 2020. 'YouTubers unite': collective action by YouTube content creators. Transf. Eur. Rev. Labour Res. 26, 2 (2020), 223–227. DOI:https://doi.org/10.1177/1024258920920810

[69] Abby Ohlheiser. 2018. A week later, YouTube condemns a Logan Paul vlog of a suicide victim's body, says it's looking at 'further consequences.' The Washington Post. Retrieved January 7, 2021 from https://www.washingtonpost.com/news/the-intersect/wp/2018/01/09/a-week-later-youtube-condemns-a-logan-paul-vlog-of-a-suicide-victims-body-says-its-looking-at-further-consequences/

[70] Marie Page. 2017. Facebook Group Member Request Settings: new Facebook feature. The Digiterati. Retrieved January 15, 2021 from https://thedigiterati.com/facebook-group-member-request-settings/

[71] William Clyde Partin. 2020. Bit by (Twitch) Bit: "Platform Capture" and the Evolution of Digital Platforms. Soc. Media + Soc. 6, 3 (July 2020), 205630512093398. DOI:https://doi.org/10.1177/2056305120933981

[72] Patel Sahil. 2017. The "demonetized": YouTube's brand-safety crackdown has collateral damage. Digiday. Retrieved from https://digiday.com/media/advertisers-may-have-returned-to-youtube-but-creators-are-still-losing-out-on-revenue/

[73] Hector Postigo. 2016. The socio-technical architecture of digital labor: Converting play into YouTube money. New Media Soc. 18, 2 (2016), 332–349. DOI:https://doi.org/10.1177/1461444814541527

[74] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as mechanisms for supporting algorithmic transparency. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 1–13. DOI:https://doi.org/10.1145/3173574.3173677

[75] Randhawa Kay Lily. 2020. How to Navigate YouTube's Unclear Demonetization System. Superjump. Retrieved from https://medium.com/super-jump/how-to-navigate-youtubes-unclear-demonetization-system-5c437c70e0ae

[76] Noopur Raval and Paul Dourish. 2016. Standing out from the crowd: Emotional labor, body labor, and temporal labor in ridesharing. In Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW, Association for Computing Machinery, New York, NY, USA, 97–107. DOI:https://doi.org/10.1145/2818048.2820026

[77] Ivan Rivera. 2019. CRAN - Package RedditExtractoR. Retrieved from https://cran.r-project.org/web/packages/RedditExtractoR/index.html

[78] Sarah Roberts. 2016. Commercial Content Moderation: Digital Laborers' Dirty Work. Media Stud. Publ. (January 2016). Retrieved from https://ir.lib.uwo.ca/commpub/12

[79] Alex Rosenblat and Luke Stark. 2016. Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers. Int. J. Commun. 10, (2016), 3758–3784. DOI:https://doi.org/http://dx.doi.org/10.2139/ssrn.2686227

[80] Koustuv Saha and Munmun De Choudhury. 2017. Modeling stress with social media around incidents of gun violence on college campuses. Proc. ACM Human-Computer Interact. 1, CSCW (November 2017), 1–27. DOI:https://doi.org/10.1145/3134727

[81] Sarita Schoenebeck, Oliver L Haimson, and Lisa Nakamura. 2020. Drawing from justice theories to support targets of online harassment. New Media Soc. (March 2020), 146144482091312. DOI:https://doi.org/10.1177/1461444820913122

[82] Trebor Scholz. 2012. Digital Labor: The Internet as Playground and Factory. Routledge.

[83] James C. Scott. 1985. Weapons of the Weak: Everyday Forms of Peasant Resistance - James C. Scott - Google Books.

[84] Diana Secara. 2015. The Role of Social Networks in the Work of Terrorist Groups. Research and Science Today, 77–83.

[85] Joseph Seering. 2020. Reconsidering Self-Moderation: the Role of Research in Supporting Community-Based Models for Online Content Moderation. Proc. ACM Human-Computer Interact. 4, CSCW2 (October 2020), 1–28. DOI:https://doi.org/10.1145/3415178

[86] Joseph Seering, Tony Wang, Jina Yoon, and Geoff Kaufman. 2019. Moderator engagement and community development in the age of algorithms. New Media Soc. 21, 7 (July 2019), 1417–1443. DOI:https://doi.org/10.1177/1461444818821316

[87] Qinlan Shen, Michael Miller Yoder, Yohan Jo, and Carolyn P Rose. 2018. Perceptions of Censorship and Moderation Bias in Political Debate Forums. Int. AAAI Conf. Web Soc. Media; Twelfth Int. AAAI Conf. Web Soc. Media (2018). Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17809/17026

[88] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. 2016. Confessions: Grand challenges for HCI researchers. Interactions 23, 24–25. DOI:https://doi.org/10.1145/2977645

[89] Spandana Singh. 2019. Everything in Moderation: An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content. Retrieved from

https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/

[90] Emily Stewart. 2019. "We don't want to be knee-jerk": YouTube responds to Vox on its harassment policies - Vox. Vox. Retrieved from https://www.vox.com/recode/2019/6/10/18660364/vox-youtube-code-conference-susan-wojcicki-carlos-maza

[91] Cynthia Stohl, Michael Stohl, and Paul M. Leonardi. 2016. Digital Age | Managing Opacity: Information Visibility and the Paradox of Transparency in the Digital Age. Int. J. Commun. 10, (2016), 123–137. Retrieved December 28, 2020 from https://ijoc.org/index.php/ijoc/article/view/4466

[92] Nicolas Suzor, Tess Van Geelen, and Sarah Myers West. 2018. Evaluating the legitimacy of platform governance: A review of research and a shared research agenda. Int. Commun. Gaz. 80, 4 (2018), 385–400. DOI:https://doi.org/10.1177/1748048518757142

[93] 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. Int. J. Commun. 13, (2019). Retrieved from https://ijoc.org/index.php/ijoc/article/view/9736

[94] Twitter. 2015. Securities and Exchange Commission Filing. Retrieved from https://www.sec.gov/Archives/edgar/data/1418091/000156459015004890/twtr-s-3_20150605.htm

[95] 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. In Proceedings of the ACM on Human-Computer Interaction, Association for Computing Machinery, 1–22. DOI:https://doi.org/10.1145/3415238

[96] Michael A. De Vito, 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, Association for Computing Machinery, New York, NY, USA, 3163–3174. DOI:https://doi.org/10.1145/3025453.3025659

[97] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From human-human collaboration to Human-AI collaboration: Designing AI systems that can work together with people. In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2020), Association for Computing Machinery, New York, NY, USA, 1–6. DOI:https://doi.org/10.1145/3334480.3381069

[98] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing theory-driven user-centric explainable AI. In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), Association for Computing Machinery, New York, New York, USA, 1–15. DOI:https://doi.org/10.1145/3290605.3300831

[99] Michael Wenzel, Tyler G. Okimoto, Norman T. Feather, and Michael J. Platow. 2008. Retributive and restorative justice. Law and Human Behavior 32, 375–389. DOI:https://doi.org/10.1007/s10979-007-9116-6

[100] Wayne W. Wilkinson and Stephen D. Berry. 2019. Together They Are Troy and Chase: Who Supports Demonetization of Gay Content on YouTube? Psychol. Pop. Media Cult. (2019). DOI:https://doi.org/10.1037/ppm0000228

[101] Donghee Yvette Wohn. 2019. Volunteer moderators in twitch micro communities: How they get involved, the roles they play, and the emotional labor they experience. In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), Association for Computing Machinery, New York, New York, USA, 1–13. DOI:https://doi.org/10.1145/3290605.3300390

[102] Donghee Yvette Wohn, Guo Freeman, and Caitlin McLaughlin. 2018. Explaining viewers' emotional, instrumental, and financial support provision for live streamers. Proc. 2018 CHI Conf. Hum. Factors Comput. Syst. 2018-April, (2018), 1–13. DOI:https://doi.org/10.1145/3173574.3174048

[103] Lindsey Wotanis and Laurie McMillan. 2014. Performing Gender on YouTube. Fem. Media Stud. 14, 6 (November 2014), 912–928. DOI:https://doi.org/10.1080/14680777.2014.882373

[104] Eva Yiwei Wu, Emily Pedersen, and Niloufar Salehi. 2019. Agent, gatekeeper, drug dealer: How content creators craft algorithmic personas. Proc. ACM Human-Computer Interact. 3, CSCW (2019), 1–27. DOI:https://doi.org/10.1145/3359321

[105] Katrina Wu. 2016. YouTube Marketing: Legality of Sponsorship and Endorsement in Advertising. J. Law, Bus. Ethics 22, (2016), 59.

[106] Keng Chieh Yang, Chia Hui Huang, Conna Yang, and Su Yu Yang. 2017. Consumer attitudes toward online video advertisement: YouTube as a platform. Kybernetes 46, 5 (2017), 840–853. DOI:https://doi.org/10.1108/K-03-2016-0038

Received January 2021; revised April 2021; accepted July 2021.