# Peer Produced Friction: How Page Protection on Wikipedia Affects Editor Engagement and Concentration

LEAH AJMANI, University of Minnesota, USA
NICHOLAS VINCENT, University of California, Davis, USA
STEVIE CHANCELLOR, University of Minnesota, USA

Peer production systems have frictions–mechanisms that make contributing more effortful–to prevent vandalism and protect information quality. Page protection on Wikipedia is a mechanism where the platform's core values conflict, but there is little quantitative work to ground deliberation. In this paper, we empirically explore the consequences of page protection on Internet Culture articles on Wikipedia (6,264 articles, 108 edit-protected). We first qualitatively analyzed 150 requests for page protection, finding that page protection is motivated by an article's (1) activity, (2) topic area, and (3) visibility. These findings informed a matching approach to compare protected pages and similar unprotected articles. We quantitatively evaluate the differences between protected and unprotected pages across two dimensions: editor engagement and contributor concentration. Protected articles show different trends in editor engagement and equity amongst contributors, affecting the overall disparity in the population. We discuss the role of friction in online platforms, new ways to measure it, and future work.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; **Wikis**.

Additional Key Words and Phrases: Social Media/Online Communities; Empirical study that tells us about how people use a system; Quantitative Methods

**ACM Reference Format:**
Leah Ajmani, Nicholas Vincent, and Stevie Chancellor. 2023. Peer Produced Friction: How Page Protection on Wikipedia Affects Editor Engagement and Concentration. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 349 (October 2023), 33 pages. https://doi.org/10.1145/3610198



## 1 INTRODUCTION

In the design of user-centered systems, there is a constant tension between facilitating and limiting user interactions. Frictions, changes in interaction to make systems more taxing to use [32], have long been a core interest of CSCW and HCI researchers [22, 57, 102]. Frictions have been studied in peer production [30, 40], social platforms [16, 17, 47], and other systems [2]. In addition to limiting interaction by certain groups, frictions can be leveraged as lightweight interventions with different values at stake: to combat poor peer citizenship, encourage quality content [31], and temper misinformation spread [35, 46]. In 2021, former Facebook employee Frances Haugen implored the United States Congress to consider "selective frictions" [71] such as Twitter's redesign during the 2020 election [28] as a much more reasonable method to combat misinformation on the platform.

Authors' addresses: Leah Ajmani, University of Minnesota, Minneapolis, USA, ajman004@umn.edu; Nicholas Vincent, University of California, Davis, Davis, USA, nmvincent@ucdavis.edu; Stevie Chancellor, University of Minnesota, Minneapolis, USA, steviec@umn.edu.


2573-0142/2023/10-ART349 $15.00
https://doi.org/10.1145/3610198

Friction in peer production systems introduces a tradeoff: frictions often limit the participation and engagement of users to encourage platform integrity. Peer production systems, like Wikipedia, OpenStreetMap, and Linux, rely on the participation and contributions of the public to be successful [8, 9]. However, these platforms limit participation through frictions, such as requiring pseudonymous account registration, barring low-quality contributions, and peer review before publication. For Wikipedia specifically, which is the topic of our paper, maintaining platform integrity through peer government is a core value of the platform [85], and so is open participation and equality among contributors [70]. This is a deliberate decision but results in a values conflict that Wikipedia must balance.

Participation is the crux of Wikipedia's ongoing and future success; consequently, it is crucial to understand the empirical consequences of friction to ensure they have the desired effects on participation and the platform's values. Previous work has shown that platform-driven intervention (such as content moderation) can have mixed effects on the user populations they target, some positive and intentional [17, 47] and some more surprising [16]. Many of these interventions, such as hashtag banning, are frictional: they make contributing more effortful but don't block it altogether. Wikipedia is an example of a platform where the consequences of friction have significant impacts on participation, such as scaring away newcomers [40] and minority editors [27, 59]. In fact, the Wikimedia Foundation has made it a priority to onboard more diverse editors to the platform [59, 72]. Conversations about what to *do* with frictions are incomplete without exploring the *consequences* of those frictions, the disparate impacts of said frictions, and their impacts on core platform values, such as engagement and equality. As many platforms, especially Wikipedia, rely on friction-based interventions like these, understanding what happens when these frictions are deployed is crucial to facilitating constructive discussions about what they do and whether they should be used.

In this paper, we ask: *what are the motivations underlying page protection, and when page protection is deployed, are the consequences to participation and parity of editing in line with Wikipedia's goals?* Page protection is an intervention on Wikipedia that "locks" an article from editors who are not confirmed users and have limited prior edit history (<10 edits). Unlike participant banning [30] or edit reverting [40], page protection is a sweeping mechanism for a page and all its editors, not an individual. Page protection is instantiated by peers and discursive; any editor on Wikipedia can request that a page be protected, and thereby open a dialogue about protecting said article. It targets broad swathes of "bad actors" rather than single instances of rebuking poor citizenship. We argue that page protection on Wikipedia is a salient example of platform values coming into conflict through this friction and serves as an excellent opportunity to empirically understand how friction like this influences participation, especially of newcomers to the site.

To explore these tensions around friction and its outcomes on participation in peer production, we focus on the following research questions:

**RQ1. What motivates editor requests for page protection?**

**RQ2. Does editor participation change consistently and predictably when page protection is employed?**

**RQ3. How does the concentration of editing contributions change after page protections?**

Our observational study examines a particularly participatory part of Wikipedia - articles in the Internet Culture category. The Internet Culture category on Wikipedia has 6,264 articles (108 are page protected) with over 1.6 million edits. Internet Culture articles cover phenomena on the Internet – and more editors (especially newcomers) can make contributions given the lower barriers to technical knowledge. These Wikipedia articles have historically become battlegrounds for high-profile Internet controversies, such as the GamerGate harassment campaign [41, 66, 83], and are prevalent throughout the Internet ecosystem as they appear in search engines [99] and on

Twitter [118]. These distinctive participation dynamics can lead to policy conflicts and, therefore, impact participation.

We adopt a mixed-methods approach to studying motivations for and observing the consequences of page protection in Internet Culture articles. We answer RQ1 with a thematic analysis of 150 requests for page protection in the Internet Culture category. The results from RQ1 inform our variable selection and quantitative strategy for the rest of the paper. To support our observational analyses, we use a 1-1 matching method to identify pages that were protected and those that were comparable matches but ultimately were never protected. We explore the consistency and concentration of editing on Wikipedia through observational quantitative analyses in RQ2 and RQ3.

Aligning with prior work [42], we find that requests for page protection are based on a combination of the article's activity, topic, and visibility. For participation (RQ2), page protection increases *editor churn*, with dramatic increases in both user dropoff after protection and new user uptake, which runs counter to Wikipedia's goal of causing minimal damage to editor participation on protected pages. Furthermore, when pages are protected, we do not see predictable or consistent changes in editor engagement. For editor concentration (RQ3), page protection has complex interactions with editing parity. On one hand, the heaviest editors do not necessarily dominate editing after protection. However, page protection significantly increases the inequality of editing share across articles. Our results suggest that about 85% of post-protection edits would need to be redistributed to achieve perfect editor equality, compared to nearly 30% of edits on unprotected pages.

Our results suggest that protected articles experience inconsistent changes to the editor landscape and may not meet the self-stated policy goals of the community for Internet Culture articles. The heterogeneous impacts of page protection raise important questions for Wikipedians on what the tradeoffs are for enacting this "peer produced friction" and whether the bluntness of page protection is appropriate for the platform's goals. We highlight the need for scholars to consider more nuanced facets of "friction," such as the amount of damage caused and affected subjects/users. We propose more participation in deliberations about friction to combat these impacts and improve engagement in peer production systems.

## 2 RELATED WORK

### 2.1 Background

Peer production systems are when large groups of people self-organize to produce a shared outcome, whether a good or service. In CSCW and HCI, peer production and collaboration is a long-standing area of research interest on platforms like Linux, Wikipedia, and Mozilla [8, 80]. Peer production systems navigate the tension of promoting contributor engagement while maintaining the quality of the final product, whether that be software or an encyclopedia. One technique to negotiate this is to employ frictions, which intentionally makes participation more effortful.

In this section, we discuss the rich history of design frictions in socio-technical systems and how those frictions have manifested in content moderation design. Finally, we discuss how Wikipedia specifically has incorporated design frictions into its self-governing moderation strategy.

### 2.2 Friction and Participant Engagement in Social Systems

The term "design friction" often refers to pain points during an interaction between a user and a technology [75]. Previous work has focused on decreasing design frictions to increase usability and reduce user frustration [60]. However, optimizing for usability can cause thoughtless behavior that may increase the spread of misinformation [5, 46]. Frictions – changes to interaction to make it more taxing in some way [32] – offer an interesting solution to the over-optimization of usability.

Frictions are already commonplace to prevent misuse. Recent work has explored creating these moments of reflection to moderate the flow of information on peer production sites [36, 46]. For example, pop-up dialogues have been employed to warn users of security issues [24], causing a moment for users to stop and think about whether they want to proceed. Frictions can also be helpful to combat bad actors from creating addictive systems through dark patterns [33]. Some designers have even advocated for slow design [36, 89], a method of actively designing for slower interactions that promote intention and reflection.

However, recent work has shown that users can become habituated to minor interruptions and frictions [2], potentially rendering them ineffective. Even design affordances built to combat misinformation, such as Facebook's flagging feature, have been shown to backfire and make the flagged content more salient in the interface [1]. To combat misinformation surrounding the 2020 election, Twitter started placing warnings over potentially misleading information where users have to actively dismiss the notification to then see the content [28]. Jahanbakhsh et al. [46] found that behavioral nudges, such as providing an accuracy assessment or rationale of a claim, reduced the sharing of false content. However, these nudges decreased overall sharing, thereby, reducing the sharing of true content as well.

Our work builds off this exploration of tradeoffs by exploring the consequences of an intervention where the self-stated goal is to block certain groups of editors. Specifically, we explore whether the impacts of Wikipedia page protection on editor engagement and editor equality are consistent and, therefore, foreseeable.

### 2.3 Moderator Interventions and Tradeoffs

In HCI research, frictions are often studied in the context of moderator interventions. Moderation is a friction that deals with the tradeoff of quality vs. participation. This is relevant to peer production because of page protection's similar effect of preventing some people from participating on a page. In this section, we overview related work on moderation and its impacts on communities.

Grimmelmann [34] defines content moderation as the "the governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse," noting that maintaining the balanced tension between community participation and abuse prevention is essential to the goals of moderator interventions. This balance implies that moderator interventions should hold two components: (1) effectively prevent abuse and (2) facilitate engagement and cooperation. In HCI, research has focused on many goals encompassing this idea of "abusive behavior," from undesirable content (like spam and unwanted pornography) to anti-social behaviors [19]. This includes reducing hate speech [18], limiting the spread of misinformation [46], and preventing vandalism [30]. However, moderators also consider the tradeoff of facilitating engagement and cooperation. Recent work has focused on proactive and positive moderation interventions [37]. For example, Seering et al. [87] note there are many nurturing moderator roles, such as curation or gardening. Gilbert [31] found similar actions in r/AskHistorians.

While page protection is a mechanism specific to Wikipedia, it parallels many other blocking mechanisms on peer production platforms. Many platforms employ administrator techniques to block or alter user-generated content, such as restricting public access to content [17] or tagging it to inform users of potentially taboo topics [26]. For example, Reddit moderators can lock threads as they see fit [62]. The New York Times automatically closes the comments section on articles 24 hours after publication [68]. Due to challenges around vandalism, poor citizenship, and information quality, peer production systems often have mechanisms to limit who can contribute information and how. On Wikipedia specifically, previous work has explored the unintended consequences of anti-vandal mechanisms [30] and how the larger goal of quality control impacts the peer production ecosystem [92].

Most closely related to our work is the group of quantitative research that has evaluated the effectiveness of moderation in meeting the self-stated goals of platforms. For example, Chandrasekharan et al. [17] found that banning hateful subreddits limits hate speech on Reddit. Likewise, Jhaver et al. [47] found that deplatforming controversial figures, such as Alex Jones, effectively tempered conversations about them and limited toxicity amongst their supporters. On the other hand, work by Chancellor et al. [16] found that users subverted content moderation of dangerous disordered eating hashtags, and moderation may have increased the severity of the content. In the context of self-governance, Fan and Zhang [25] evaluated the effectiveness of a civic jury model in deliberating digital governance systems and increasing perceptions of justness.

In our work, we focus on a moderator intervention that creates friction by making a contribution more effortful for the user. Furthermore, we build on the evidence from content moderation studies to explore how page protection on Wikipedia may positively or negatively impact user engagement.

## 2.4 Wikipedia, Friction, and Page Protection

Wikipedia is a peer production platform famous for allowing anyone to edit it [27, 70]. Due to Wikipedia's open contribution model, efforts to preserve information quality on Wikipedia are essential to the platform's success [84]. Therefore, Wikipedia has numerous mechanisms to limit the effects of bad actors. For example, Wikipedia administrators can ban specific usernames [30], revert damaging edits [40], and protect pages [42]. Many solutions come in several affordances and forms. For example, preventing misinformation on Wikipedia uses human fact-checking and bot intervention. However, human fact-checking is resource intensive [68], and bots often miss nuance and context [96].

Page protection on Wikipedia is an intervention that administrators can impose on a page and limits who can change the page through edits, moves, creation, or similar actions. Moreover, page protection limits participation by blocking users with low credentials, such as anonymous users or with low edit history. In theory, this implies that an individual can "overcome" the block by verifying themselves or editing other pages. While any user can request page protection, only administrators can apply protection to a page. The Wikipedia policy states that:

> "While Wikipedia strives to be as open as possible, sometimes it is necessary to limit editing of certain pages in order to prevent vandalism, edit warring, or other disruptive edits." [105]

The self-stated policies around page protection are relatively straightforward, noting that page protection runs counter to Wikipedia's open mission, and preemptive protection is generally not allowed. In other words, the Wikipedia policy states that page protection is a frictional mechanism but should be used only after evidence of edit warring, vandalism, or other disruptive behavior.

Per the Wikipedia policy, page protection is typically applied on pages used as venues for poor Wikipedian citizenship. Page protection is exciting, given its relationship to high-profile articles. High-profile articles are often the targets of "bad actors" and are often page protected. For example, the main page of Wikipedia has been protected since 2006 [42]. Previous work has mainly focused on detecting page protection of Wikipedia [42, 91] and painting broad portraits of the intervention. For example, Hill and Shaw [42] highlight the pervasive impact of page protection on the platform. They note that protected pages are often heavier viewed and, therefore, more influential. We extend this work by focusing on page protection's impact on the editor landscape at the article level.

Our work bridges these three areas by focusing on page protection's social consequences. Page protection balances these core tensions in peer production and content moderation about the openness of contribution and preserving quality. However, without understanding the consequences of page protection, page protection contributes to Wikipedia's "hidden order" and bureaucratic

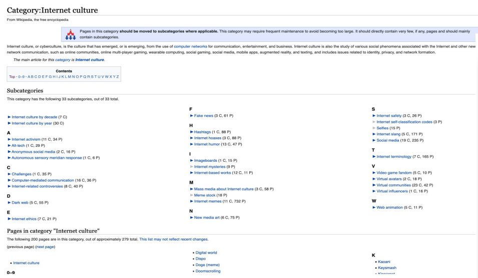


Fig. 1. Wikipedia category page for "Internet Culture." A category page on Wikipedia can link to other pages and subcategories, creating a tree-like structure with pages.


Fig. 2. Example of an edit page protected article with a "blue lock" (circled in red).

nature [14, 82, 88, 98]. Our study, therefore, evaluates the consequences of page protection on editor participation and engagement in articles.

## 3 METHODS OVERVIEW

In our observational study of page protection, we employ qualitative and quantitative methods to understand the *motivations* and *consequences* behind the intervention. To better understand the motivations for page protection and to identify latent variables for our computational analysis, we use thematic analysis [21] on the requests for page protection. Similar to prior work in CSCW and Wikipedia [44], these results directly informed our quantitative work where we employ matching methods to create groups of protected and unprotected pages.

Matching methods in observational is a technique borrowed from the quasi-causal methods space and is especially useful in social computing research where running our experiments would be infeasible or, especially in our case, very unwanted or unethical (similar to Chandrasekharan et al. [18]). Prior work in HCI / CSCW to has used these techniques with observational data [45, 79, 95] to study many areas, such as measuring mental health [23], and the impact of banning on communities [18]. In the next few sections, we describe our dataset, our qualitative methods and results, and how we use our qualitative results to inform quantitative methods design.

## 4 DATASET SOURCE: WIKIPEDIA & INTERNET CULTURE CATEGORY

Wikipedia pages are often placed in categories such as "Living persons" or "List of wars by country," to better organize and navigate the massive amount of content on the website. To focus on socially emerging topics with potential for broad participation, we studied articles within the Internet Culture category. Wikipedia defines Internet Culture as, "the culture that has emerged or is emerging from the use of computer networks for communication, entertainment, and business," [106] noting that social phenomena are intertwined with this culture.

The Internet Culture category has unique participation dynamics, making it an excellent candidate for our research interests. Internet culture is one of the most accessible categories for newcomers; yet, it is ripe for conflicts of interest and controversy. Internet culture does not require technical expertise, meaning lay or new editors can participate more easily. Previous work has explored scientific articles on Wikipedia as cases for an emerging topic prone to misinformation [53]. However, these articles require a certain amount of expertise and technical knowledge for substantive contributions. They also need more knowledge about Wikipedia practices, like the "neutral point of view" policy. Therefore, technical article editing naturally poses a barrier for newcomers [13, 77]. Furthermore, given that newcomers often edit articles of personal interest to them [4], we can intuit that those interested in internet culture have a level of technical accessibility and literacy that makes it easy for them to edit Wikipedia. Empirically, this ease of access to Wikipedia married with the emerging nature of topics has led to numerous controversies. For example, GamerGate, an internet harassment campaign outside of Wikipedia, led to polarizing moderator interventions within the platform [83]. Furthermore, viral events or influencers linking Wikipedia articles can affect the editor landscape.

On Wikipedia, categories follow a tree-like structure: one category can link to both pages and subcategories. For example, the "Category: Internet Culture" page links to subcategories, such as "Internet hoaxes" and articles, such as "Digital world" (See Figure 1). The "Internet hoaxes" category links to its own set of subcategories and pages. We generated our dataset with a depth-first search through all subcategories, with "Internet Culture" as the root. This method generated 6,264 articles over 20 years (2002-2022) as of the date of our data collection (Aug 8, 2022), 108 (1.7%) of which are currently edit-protected without an automatic unprotection date. There are 1,647,797 edits made by 359,969 unique user IDs. Because we looked at editors in terms of their user identifiers, different IP addresses–which is how anonymous users are represented–are considered different users.

## 5 RQ1: QUALITATIVE METHODS

To study the impacts of page protection on editor participation, we must first understand *how* page protection is discussed and enacted. We used thematic analysis [21, 44] on requests for page protection to answer RQ1: *what motivates editor requests for page protection?* Especially given the participatory nature of the Internet Culture category, a qualitative analysis of requests for page protection provides unique insights into Wikipedian's motivations behind page protection and how it fits into the larger ecosystem of information and participatory engagement in the community. Pragmatically, reasons for page protection inform our downstream matching strategy so we can create comparable sets of protected and unprotected pages (See section 7.2) to answer RQ2 and RQ3. This approach has been used in recent Wikipedia work [44] to support more rigorous quantitative analysis.

### 5.1 Dataset: Requests for Page Protection (RFPPs)

Anyone on Wikipedia can request a page be protected by submitting their request and reasoning through a specific form [115], called a request for page protection (RFPP). The request is then escalated to relevant administrators, who can decline, approve, or take other action on the request, such as blocking specific users. Wikipedia logs all requests made and relevant discussions through the RFPP archives [114]. We scraped logs to get all RFPPs made on Wikipedia articles within the Internet Culture category, leading to 462 requests for page protection.

### 5.2 Thematic Analysis

Building on Houtti et al. [44], we used thematic analysis to qualitatively study RFPPs [21]. The lead author began by randomly selecting 50 RFPPs. They open-coded the requests for rationales given or implied by the requester and why administrators accepted or declined the request. They also logged the administrator's action (e.g., semi-protection, user blocked, declined). They then clustered the codes into themes for motivations for page protection. Next, the first author randomly sampled 10 requests at a time, open-coded, and refined the themes as necessary. They reached theoretical saturation after coding 100 total requests as new codes stopped emerging. To verify saturation, the lead author continued coding and clustering 50 more requests in batches of 10. They consulted with the other authors during this process to ensure the themes were sensible and the coding was

| Article Activity | Example Quote |
|---|---|
| Vandalism | *"Persistent vandalsim"* |
| Disruptive editing | *"Persistent sock puppetry, additions of uncited content, and all around disruptive editing."* |
| IP editing | *"Three different anon IPs have all made the same undiscussed, non-consensus infobox edits"* |
| **Article Topic** | **Example Quote** |
| Biography | *"Continued restoration of negative content of a living person (WP:BLP) using an unreliable source"* |
| Emerging Topic | *"Persistent disruptive edits... Article is about an upcoming amateur boxing match"* |
| **Article Visibility** | **Example Quote** |
| Event Trigger | *"Persistent Vandalism. Viral event is drawing memers and Redditors from far and wide to disrupt the article."* |
| External Visibility | *"Page has been linked by TikTok user [anon] in a recent TikTok. This users links in the past have encouraged a wave of vandalism"* |
| Internal Visibility | *"Long term vandalism on popular page"* |

Table 1. Results from our thematic analysis of requests for page protection ($n$ = 150). We found three main criteria types: activity on the article, the topic of the article, and visibility both on Wikipedia and other platforms such as TikTok and Reddit.

appropriate. In total, we generated 150 open codes, eight main criteria for page protection, and three major themes.

## 6 RQ1 RESULTS: WHAT MOTIVATES REQUESTS FOR PAGE PROTECTION?

In this section, we present our qualitative analysis of requests for page protection (Table 1) and our resulting matching criteria. Our thematic analysis reveals several criteria Wikipedians commonly use to motivate the need for page protection. We separate these reasons into three broad categories: article activity, article topic, and article visibility.

### 6.1 Article activity

Recall that page protection is a reactive measure rather than a preemptive one. Articles often experience disruptive activity that causes a reactive request for page protection. Below, we describe a few of these reasons.

**Vandalism.** Vandalism was the most commonly cited reason in the requests for page protection. Often, the request was short and to the point – "*Persistent vandalism*," whereas other requests would describe the vandalism, such as "*Persistent vandalism – IP vandalisms, especially mass blanking, all day*". Some requests cited previous ways editors attempted to handle the vandals, such as the three-revert-rule [104], which states that an editor must not perform more than three edit reverts on a single page in 24 hours– "*Please also note I'm at 3RR so the page is currently in the bad state*."

**Disruptive editing.** Disruptive editing on Wikipedia is an edit activity pattern that prevents the improvement of an article. Wikipedia notes that this is distinct from vandalism – "Disruptive editing is not always vandalism, though vandalism is always disruptive" [110]. Disruptive editing is often invoked with other reasons to describe poor citizenship. For example, "*Persistent sock puppetry, additions of uncited content, and all around disruptive editing.*" Disruptive editing is often used when the requester cannot assess the intentions of the policy violators. For example, some requests will note the actions without calling the user a vandal. "*Persistent disruptive editing – IP hopper making unnecessary changes to the infobox*." Furthermore, we found that IP editors were often mentioned as culprits of disruptive editing.

**IP Editing.** IP editing is when users without formal Wikipedia accounts edit an article, which makes their IP address their username. Although IP editing enables open contribution, IP editors were listed as a common cause of vandalism or disruptive editing. Some requests cited groups of IPs, "*three different anon IPs have all made the same undiscussed, non-consensus infobox edits*," while others noted their suspicions of an IP-hopper (a single person using multiple IPs), "*strongly suggesting it is* *the same person using different IPs to edit-war*." Disruptive IP editors can often be handled through a mechanism called "range blocking," which only bans a specific range of IP addresses instead of banning all IP editors through page protection [112]. However, some requests noted that range blocking would not be the most constructive solution, "*users are IP hopping and successfully evading imposed blocks - a range block would probably take too long to calculate and impose*."

## 6.2 Article Topic

The next theme in our analysis of RFPPs was article topic.Wikipedia policy states that disputes about an article's content should be resolved through discussion on talk pages rather than page protection requests [107]. However, an article's subject matter can still play a key role in its page protection.

**Biography.** The biographies of living persons (BLP) category on Wikipedia is a unique group of articles that have specific policies to protect from conflicts of interest, privacy violations, and general poor citizenship [108]. Because this topic has clear policy guidelines, some requests for page protection simply noted the policy, writing "*BLP gone crazy*" or "*BLP policy violation*." Others outlined the specific harmful actions that users were taking, "*Continued restoration of negative content of a living person (WP:BLP) using an unreliable source*."

**Emerging topic.** Major events, such as crises or celebrity deaths, often influence Wikipedia articles [51, 52]. We find that major and minor events can motivate page protection. For example, one request cited an amateur boxing match as drawing bad actors, "*Persistent disruptive edits and vandalism...Article is about an upcoming amateur boxing match*." Other requests would cite cyclical topics as reasons to protect the page. For example, an annual event in autumn was requested for page protection at the beginning of November: "*Persistent vandalism – It's that time of year again*."

## 6.3 Article Visibility

Previous work has noted that protected articles are often higher-profile within the platform [42]. For example, pages that are linked on the main page are often candidates for page protection. We find that internal popularity is a motivator for page protection, but also external links can affect pages just the same.

**Public events.** Distinct from emerging topics, we found that public events that run tangentially to the article's topic can still draw traffic and poor citizenship. For example, a viral event, *Josh Fight*, affected the article *Josh*, causing it to be a candidate for page protection, "*Persistent Vandalism. Viral event is drawing memers and Redditors from far and wide to disrupt the article. Please protect quickly*."

**External visibility.** Wikipedia articles are often linked on other platforms and are pervasive in search engines [99]. We found that external links, links to the article on platforms besides Wikipedia, were often motivators of page protection. For example, one request cited that the article had been linked on TikTok, "*Page has been linked by TikTok user [anon] in a recent TikTok. This users' links in the past have encouraged a wave of vandalism...*"

**Internal visibility.** Internal notability–an article's prominence within Wikipedia–also motivates page protection. This often was as simple as requesters noting that the page is popular, "*Long term vandalism on popular page*," while others cited internal events increasing the page's visibility. For example, one request noted that the article's deliberation was further reason to protect it, "*A discussion is underway on the talkpage but it looks as though the page isn't going to remain stable while that's concluded.*"

## 6.4 Factors that do not influence page protection decisions

There were also common themes among declined requests. Notably, requests that only mentioned the injection of false content were declined due to Wikipedia's content dispute policy [107]. For

| Theme | Matching Criteria |
|---|---|
| Article Activity | Pre-protection edits, detecting vandalism instances, IP users |
| Article Topic | Subcategory |
| Article Visibility | Pageviews |

Table 2. Matching criteria informed by our qualitative analysis. We use these measures as a proxy for whether a page could have reasonably been requested for page protection (See section 7.2).

example, one request stated: *"This page has been regularly visited by several IPs who are attempting to insert the partisan, unencyclopedic, and false prose 'It is considered a Clinton outrage machine which has set a $40 million budget to oppose President Trump.'"* This request was declined as the dispute resolution policy states that content disagreements should be discussed on talk pages rather than through interventions, such as page protection. While our findings demonstrate that the article's topic informs page protection, they also show that content disputes are beyond the scope of page protection.

### 6.5 Matching Criteria - Connecting Qualitative to Quantitative

This section describes how our qualitative findings informed our matching criteria (Table 2). By connecting the themes we discovered to metrics, we can create a set of unprotected but comparable articles. We describe how our matching criteria are calculated and used in our one-to-one matching algorithm in Section 7.2.

**Article Activity.** We found three main characteristics invoked about page protection: vandalism, edit activity, and IP users. All three metrics have clear parallels to the Wikipedia edit history we used as matching criteria: the number of vandal edits detected, the number of edits in a thirty-day period, and the number of unique IP users. As mentioned above, content disputes, such as biased sources, are resolved on talk pages and are not sufficient reasons to protect a page. Therefore, we ignored content-based metrics typically associated with an article's activity such as citation count [81] and article length [12].

**Article Topic.** Our qualitative findings show that an article's topic influences page protection insofar as it affects the policies surrounding an article. For example, biographies of living people (BLP) often have strict guidelines cited in page protection requests. Moreover, content of an article does not contribute to page protection. Therefore, we operationalize "topic" as an article's category rather than the article's body. Within the Internet Culture category we study, numerous sub-categories such as "internet personalities," have their own rules. We capture this by matching on a page's immediate sub-category, defined in Section 7.2.

**Article Visibility.** Previous work has noted that protected articles are often higher profile [42]. Our qualitative analysis demonstrates that events outside of the platform affect these specific articles, and bring attention to the page. To operationalize this, we focus on the pageview count per article, which captures traffic that both internal and external events can cause.

## 7 QUANTITATIVE METHODS

In this section, we outline the methods we used to explore RQ2-3. Inspired by quasi-causal methods [45, 79, 90], we employed a one-to-one matching process to create two comparable sets of articles: (1) a set of salient protected articles and (2) a set of articles that could have feasibly been page protected but were not. We leverage the comparisons through these methods to draw conclusions about our observational analysis of page protection and how it impacts editor engagement in the Internet Culture category.

| | Treatment (n=108) | Match (n=108) | Candidate (n=5,802) | English Wikipedia |
|---|---|---|---|---|
| Avg Protected Age | 809 days (SD = 1190.30) | X | X | X |
| Edits per Article | 1931 | 911 | 233 | 170 |
| Users per Article | 539 | 321 | 51 | 6.8 |

Table 3. Treated articles are, on average, protected about 800 days (approx. 2 years) after they're created, However, protection age has a standard deviation of 1190 days, suggesting that mean is not reflective of a standard protection age. Furthermore, we see treated (protected) articles generally have more users and edits than typical of articles in the English Language Wikipedia. English Wikipedia statistics taken from [109].

### 7.1 Identifying Our Treatment Set

First, we describe how we selected our "treatment" articles or those protected without an automatic expiration date. While there are multiple levels of page protection (also called "locks", after their visual padlock icon (Figure 2)), we focus on page protection that impacts participatory editing by managing users' editing rights and abilities. The two edit protection levels for articles are extended confirmed **[blue lock]** and semi-protected **[silver lock]**, each with the following policies [105]:

- **Blue Lock.** Extended confirmed protection, also known as 30/500 protection, only allows edits by editors with the extended confirmed user access level, granted automatically to registered users with at least 30 days tenure and more than 500 edits.
- **Silver Lock.** Semi-protected pages cannot be edited by unregistered users (IP addresses), as well as accounts that are not confirmed or autoconfirmed (accounts that are at least four days old and have made at least ten edits to Wikipedia).

We made two choices about data we chose not to include in our analysis. First, we ignored articles that only had the Green Lock (preventing page movement) because that protection level impacts who can *move* the page but does not affect who can *edit*. Second, we only analyzed articles where the page protection had no automatic expiration date to ensure we explored the most salient uses of the intervention [91] – expiring page protections are uncommon on Wikipedia. Page protection without an automatic expiration means that an unprotection request must be made to change the article's status – and therefore is a deliberative action.

We found 108 articles within the Internet Culture category that meet these criteria. We manually inspected 20 random articles from our analysis to ensure we included the correct locks in our filtering process.

We used the MediaWiki API [103] to gather data about page protection. To identify a date of page protection, we used the date indicated in the respective request for page protection, or the RFPP, as mentioned previously in Section 5. As mentioned, this process gave us 108 articles that are edit-protected without an expiration date (1.7% of the total in the Internet Culture category). As demonstrated in Table 3, these articles had an average protection age of 2 years and 539 editors per article. The Internet Culture category includes articles about memes, viral challenges, and Internet events (e.g., Killing of Harambe, Blue Whale Challenge, and 2016 Webby Awards).

### 7.2 Identifying Our Match Set

Next, we describe our matching process to identify articles not protected in the Internet Culture category, similar to prior work [45, 79, 95]. Our goal was to create a set of articles that could have been reasonable candidates for page protection but were ultimately not protected. One challenge in this analysis is that we do not have access to the complete data required to estimate a truly causal effect of page protection. Because we lack internal data, we cannot know if articles in our match set were indeed on the verge of being protected more than what is implied in the RFPP. This also limits our ability to leverage truly quasi-causal methods, such as Interrupted Time Series or Regression Discontinuity Design methods [55].

To create the most comparable match set, we use a one-to-one matching strategy [90] rather than random selection, which helps account for any inherent differences between articles that undergo page protection versus those that do not. Matching strategies have been demonstrated to be helpful in similar evaluations of platform frictions, like banning [18] and content moderation [16]. Matching facilitates more directed and contextualized comparisons than if we examined page-protected articles compared to all other Wikipedia articles or just page-protected articles by themselves (with no comparison point).

*7.2.1 Variables of Interest.* We operationalize the matching metrics generated from two sources: prior work on Wikipedia and our qualitative analysis (Table 2) as features in vectors for each treated article. Recall that the treatment set ($n$ = 108) is the subset of Internet Culture articles ($n$ = 6,264) that had edit-based protections without an automatic expiration date. Given an article $a$, its matching criteria vector $v_a$ contains the following:

- **category**: For a given treated article, the set of potential matches is constrained by the set of articles within the same *immediate parent category* (See Section 4). Because specific topics may be more prone to engagement, controversy, and page protection, we used subcategories as a matching criterion to control for the effects of topic on behavior (Table 2). For example, our dataset includes connected categories, such as "Internet memes," and more obscure ones, such as "Interactive artists."
- **active_age**: the time between the creation of $a$ and the time it was last edited. This feature allows us to match articles with the same time range for potential editing activity. For example, it would not make sense to compare the editing activity of an article that has existed for one day versus one year because the latter has had more time for potential edits and controversies. Formally, $active\_age(a) = date\_last\_edited(a) - date\_created(a)$.
- **edits**: the number of edits in the thirty days before $a$ was protected. For unprotected articles, this is the number of edits thirty days before the *faux protection date*, defined below. The editing rate helps capture the potential for disruptive editing, which we found to be a common reason for requesting page protection (See Section 6). Previous work has also explored how editing rate affects an article's potential to be page protected [91].
- **pageviews**: the number of pageviews of $a$ in the last thirty days of its activity. This feature captures an article's "visibility" on the platform as protected articles are often higher profile within the platform [42]. Pageviews also capture traffic due to external circumstances that we found in our qualitative analysis.
- **detected vandalism instances**: the number of edits on $a$ detected as vandalism. We use ClueBot for vandalism detection, a common benchmark in prior work [29, 58, 100, 111]. Our qualitative analysis found that vandalism is a common reason for requesting page protection, especially when other anti-vandalism techniques, such as reverts [40] or bots [30] are not alleviating the problem. To only consider vandalism instances that contributed to the request for page protection, we counted the total number of ClueBot edits before the article's protection date.
- **anonymous users**: the number of unique (IP) addresses that have edited $a$. This is distinct from the number of people who have edited $a$, given that a person can theoretically edit from multiple IP addresses [113]. We found that IP editors are often the culprits of vandalism or disruptive editing in requests for page protection. This is consistent with previous work on anonymous users within Wikipedia [43, 54, 91]. To focus on anonymous users that contributed to page protection reasoning, we counted the total number of anonymous users before the article's protection date. Note that for treated articles, where anonymous users cannot edit

the page after it has been protected, this is equivalent to the overall number of anonymous users.

*7.2.2 Faux Protection Date.* To facilitate this matching process, we created a *faux protection date* for each (treated article $t$, candidate article $c$) pair. Recall that potential candidate articles did not receive page protection, nor were they "randomly assigned" to a group of articles that could never receive page protection due to the observational nature of the data (and inappropriateness of this intervention). We adopted an approach very similar to Chandrasekharan et al. [17]'s study of the causal effect of quarantines on Reddit. In that work, the authors match subreddits *likely* to be quarantined to subreddits that were *actually* quarantined, collecting and comparing data around the date of quarantine. Through this process, the authors simulated what a randomized assignment of quarantining might look like through this process.

We created a *faux protection date* based on the *protection age* ($date\_protected - date\_created$) of the treated article. Namely, given a treated article $t$ and candidate article $c$, $faux\_protection\_date(t, c) = date\_created_c + protected\_age_t$.

*7.2.3 Matching Algorithm.* We then ran a matching algorithm, diagrammed in Algorithm 1. First, we separated articles into the *Treatment Set*, articles that have been indefinitely protected ($n$ = 108) and *Candidate Set*, articles that have never been requested for page protection ($n$ = 5,802). We then create a matching-criteria vector for every article in our treatment set, denoted by $v_t$. We iterate through every candidate article in the same category as the respective treated article. For every (treated article $t$, candidate article $c$) pair, we create a *faux protection date* based on the *protection age* ($date\_protected - date\_created$) of the treated article.

**Algorithm 1:** 1-1 Matching Algorithm

```
Result: Match set of unprotected articles
Candidate ← set of Internet Culture articles that have never been protected;
Treatment ← set of Internet Culture articles that are indefinitely protected;
T ← [v_t, ∀t ∈ Treatment];
for v_t ∈ T do
    for c ∈ Candidate do
        if category(c) = category(t) then
            v_ct ← matching criteria vector for c with faux protection date based on article t
            similarity(c, t) ← cos(v_t, v_ct)
        end
    end
    match_t ← c : max_{c∈Candidate} similarity(c, t)
end
```

Based on this faux protection date, we created a matching-criteria vector $v_{ct}$ for the treatment-candidate pair. For a treated article $t$, we find the most similar candidate article (a match article) by optimizing for the cosine similarity between $v_t$ and $v_{ct}$ where $c \in \{Candidate\ Articles\}$.

*7.2.4 Validation of Matching Approach.* Our process resulted in a match-set of 108 articles. The Standardized Mean Difference (SMD) is a commonly used statistic to check if a match set is balanced relative to the treatment set [3, 90, 119]. It is computed for each covariate by dividing the difference in mean value between the groups by the pooled standard deviation. SMD helps make comparisons independent of the sample size of different units used to measure different variables [3]. Here, we

| | Treatment | Declined Requests |
|---|---|---|
| **Active Age** | 0.332 | 0.067 |
| **Pageviews** | 0.466 | 0.519 |
| **Vandalism** | 0.386 | 0.165 |
| **IP Users** | 0.229 | 0.061 |
| **Pre-protection Edits** | 0.431 | XX |

Table 4. Standardized means difference tests between our match set and (1) treatment set ($n$ = 108) and (2) set of declined request pages ($n$ = 62). Because declined requests don't have "protection dates," we can't compare the number of edits in the thirty days of pre-protection.

used the SMD to ensure that our comparison set is providing truly relevant context (Table 4). If our match set were highly imbalanced – as determined by an SMD – the match set would provide less useful context.

Our results show that, while not similar enough to make causal claims, our match set provides relevant and comparable context to our treatment set. We see that our match set is balanced ($smd \leq .5$) in comparison to protected articles [119]. Furthermore, we see an even higher balance ($smd \leq .1$) between our match-set to the set of Internet Culture articles whose requests for page protection were denied ($n$=62) [55, 90]. This comparison implies that the articles in our match set meet the quantitative criteria for a page protection request. These comparisons help ground our claim that the match-set is a set of articles that could have been requested for page protection and reasonably work around the limitations of publicly available Wikipedia data.

### 7.3 Quantitative Measures

Next, we describe how we operationalize RQ2 and RQ3, our research questions around editor engagement and contributor concentration. Recall that we are interested in the participatory impacts of page protection, so we study the trajectory of pages *before* and *after* protection. For both research questions, we focused on the short to mid-term effects of the intervention to avoid the natural "burstiness" that Wikipedia articles experience over long periods of time [93, 120]. We analyzed activity across three main time frames: seven days, fourteen days, and thirty days.

| Metric | Description | Operationalization |
|---|---|---|
| **Dropoff** | How many users edited the article before page protection but not after? | $\lvert Before \setminus After \rvert$ |
| **Uptake** | How many users edited the article after page protection but not before? | $\lvert After \setminus Before \rvert$ |
| **Retention** | How many users edited the article both before page protection and after? | $\lvert Before \cap After \rvert$ |
| **Difference** | Number of unique usernames who edited before page protection minus after | $\lvert Before \rvert$ - $\lvert After \rvert$ |

Table 5. Metrics calculated for in response to RQ1. $Before$ is the set of unique user identifiers before page protection where $\lvert Before \rvert$ represents the number of items (i.e., users) in the set. $After$ is the set of unique user identifiers after page protection.

*7.3.1 RQ2. Editor Engagement.* Our second research question focuses on editor engagement and contributions to articles. Drawing on the literature on Wikipedia engagement [39, 73], we conceptualize engagement as the number of unique editor usernames and IDs that have edited a page in a given time frame. We represent this *Before* and *After* protection, where *Before* represents the unique set of all editor usernames that were edited before page protection or the faux-protection date (for articles in our match set). Likewise, *After* represents the unique set of all editor usernames that are edited after page protection. Both sets are bounded to a given time window. So, "seven days" encompasses the set of unique editors in the seven days immediately before treatment (*Before*) and the seven days after treatment (*After*). Measuring engagement in terms of time windows allows us to control for editing rate immediately surrounding protection.

Drawing again on prior work [39, 73], we operationalize the idea of editor engagement through four measures: dropoff, uptake, retention, and difference (see Table 5 for definitions and set formulas).

- *Dropoff.* The number of editors who contributed before the intervention but not after. These editors engage with the article before the intervention but are not observed after.
- *Uptake.* The number of editors who contributed after the intervention but not before. These editors engage with the article only after the protection has occurred.
- *Retention.* The number of editors who contributed both before and after the intervention, or the intersection of the two sets of users. These editors were retained throughout the intervention.
- *Difference.* The number of editors who contributed before the intervention minus the number who contributed after. A negative difference implies more people edited after the intervention than before.

It is important to note that these metrics are not linearly dependent on one another, given that they are based on unique user sets. For example, the set of users who contributed before page protection, $Before$, could contain users 1-20. Meanwhile, the set of users who contributed after, $After$, could contain users 21-40. Because $Before$ and $After$ have no intersecting users, but there was still high user movement, we would see a dropoff of 20 and an uptake of 20, but retention and difference of 0. In normalized terms, this means 100% of pre-intervention users did not contribute after, and 100% of post-intervention users were new to the landscape. In contrast, if $After$ contained users 1-40, then we would still see an uptake of 20, but a dropoff of 0. In terms of percentages, 100% of pre-intervention users were retained, while 50% of post-intervention users were new.

We calculated these metrics for each article across three different time windows: seven, fourteen, and thirty days, where a time window is $n$ days before and after protection. To contextualize our results about the experience of protected pages, we run $t$-tests between the treatment and match sets for each metric and time window.

| Index | Representation | Range | Reasoning |
|---|---|---|---|
| **Hoover Index** | proportion of all income that would have to be redistributed to achieve a state of perfect equality | [0,1] where 0 is perfect equality | Page protection is redistributive – imagine a world in which editing opportunities were redistributed amongst the population of Wikipedia editors |
| **Cumulative 20:20 Ratio** | ratio of wealth between the top 20% and the bottom 20% of a given population | [0, inf] where 1 is perfect equality | Insights into where the disparity is coming from in a given population |

Table 6. In response to RQ3, we calculated two inequality measures. The Hoover index measures equality in terms of resource redistribution. The cumulative 20:20 ratio allows us to measure disparity between heavy and light editors.

*7.3.2 RQ3. Editing Concentration.* Our third research question focuses on editing concentration amongst editors, or the relative influence a given editor has on an article in our dataset. Previous work has explored inequality on Wikipedia in terms of editor demographics [69] and roles of power users [76]. An extreme distribution of power manifests as significantly unequal contributions on Wikipedia [76].

In this paper, we build off of Ortega et al. [76] to operationalize editing parity as the differences in editing quantity to a given page in our dataset. This is a strong proxy for editing parity because the Wikipedia editor community maintains norms around the scope of a single edit. Wikipedia editors are encouraged to avoid making excessively trivial edits, or excessively large edits. Thus, while two edits may differ in terms of contribution size (e.g., a typo fix versus a detailed summarized claim from a primary source), when measuring across many articles, edit counts still provide meaningful ways to infer how much focus a given editor gives to a page.

To measure parity or inequality between editors, we draw from income inequality metrics that are often used to measure inequality in Wikipedia research [76] and other quantitative investigations of inequality in social computing [15, 62]. We focus on two income inequality metrics: the Hoover index and the 20:20 ratio (see Table 6).

- *Hoover index*: The Hoover index similarly measures inequality, but in a more interpretable fashion: a distribution's Hoover index is equal to the fraction of total resources that need to be redistributed to achieve a uniform distribution [63, 94]. We focus on this as page protection policy can be viewed as a "redistributive intervention", intended to redistribute editing opportunities by requiring that all contributors have an edit history.
- *Cumulative 20:20 Ratio*: Given Wikipedia editing is prone to a power distribution [56, 77], we also measured the 20:20 ratio – the ratio of edits contributed by top 20 percentile editors to bottom 20 percentile editors – to specifically interrogate what happens to editors at the extremes. The 20:20 ratio is undefined when the bottom 20 percentile has zero contributions, so to get article-level 20:20 ratios we looked at the cumulative edit count per user, to guarantee the ratio would always be defined.

Like RQ2, we calculate these statistics based on time (7, 14, and 30 days pre and post-protection). However, because we do not need pre and post-intervention user sets to calculate each metric, we focus on just the $n$ days post-protection, rather than a time window. We create an editor share vector $v_n$ for each time condition.

Inequality tests typically raise concerns about eligibility [94]. For example, income inequality can be skewed heavily if minors are included, because they do not have jobs in the same way as adults. To avoid this bias, we identify a contributor set in post-intervention conditions ($n = 7, 14, 30$) of all the editors who contributed pre-intervention and $n$ days after. Editors who did not contribute post-intervention are represented as 0 in the respective edit share vector. We then calculated the inequality metrics on $v_n$ per article per time condition. We ran significance tests of means ($t$-tests) comparing the inequality metrics in the pre-intervention window and all three post-interventions windows.

### 7.4 Robustness Check

We deliberately chose to focus on Internet Culture because the subjects of the articles and new editors are uniquely situated as potential stakeholders as they are already engaging with internet content and culture. However, we recognize that our dataset is not comprehensive of all Wikipedia articles. We address our small sample size by performing a robustness check to evaluate if our findings hold in another category with similar characteristics to ours.

We chose articles within the Ongoing Conflicts category on Wikipedia. This category of articles contains articles describing emerging hostilities between or within countries. Ongoing crises are another example of common emerging topics on Wikipedia [64]. Additionally, a significant portion of Ongoing Conflicts articles (16%) are protected. Dataset details and results of our robustness check can be found in the Appendix. The results are consistent with our findings and, in some cases, are an even clearer indication of the trends we discovered.

## 8 QUANTITATIVE FINDINGS

We find that page protection instantiates inconsistent frictions across articles – it is hard to evaluate what outcome protecting a page will have on the editor population. Additionally, page protection causes the concentration of contributions to increase (i.e., become less equal) overall but does not substantively increase the disparity between heavy and light editors.

| | | Dropoff | | Uptake | | Retention | | Difference | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Conditions | | Mean(SD) | Max | Mean(SD) | Max | Mean(SD) | Max | Mean(SD) | Min | Max |
| seven | match | 2.16(6.24) | 57 | 2.42(7.57) | 60 | 0.22(0.81) | 7 | -0.25(3.42) | -21 | 10 |
| | protected | 13.84(26.52) | 243 | 18.21(43.18) | 309 | 2.14(4.85) | 60 | -4.37(45.21) | -309 | 111 |
| | | **p-val = 1.37E-05*** | | **p-val = 2.35E-04*** | | **p-val = 6.69E-05*** | | **p-val = 0.35** | | |
| fourteen | match | 3.46(8.66) | 67 | 3.75(10.75) | 78 | 0.4(1.21) | 8 | -0.29(5.87) | -35 | 20 |
| | protected | 20.42(29.57) | 240 | 26.02(51.61) | 353 | 3.00(.265) | 25 | -5.59(52.68) | -353 | 85 |
| | | **p-val = 3.59E-08*** | | **p-val = 1.80E-05*** | | **p-val = 1.16E-06*** | | **p-val = 0.30** | | |
| thirty | match | 6.68(17.71) | 144 | 6.64(15.36) | 91 | 0.65(1.8) | 14 | 0.03(17.14) | -82 | 143 |
| | protected | 29.53(33.31) | 235 | 37.54(61.88) | 370 | 4.13(5.93) | 25 | -8.00(62.22) | -370 | 86 |
| | | **p-val = 1.80E-09*** | | **p-val = 1.02E-06*** | | **p-val = 2.15E-08*** | | **p-val = 0.20** | | |

Table 7. Protected articles experience significantly higher amounts of dropoff, uptake and retention. Both dropoff and uptake have relatively high means, suggesting that protected pages experience unusually high *user churn*. Significance codes: p-value < 0.05 '∗'

### 8.1 RQ2. Does editor participation change consistently when page protection is employed?

In this section, we describe friction outcomes in terms of four metrics of editor engagement: dropoff, uptake, retention, and difference. These metrics are typically associated with measuring editor rise and decline on platforms [39] or post-intervention [20]. As stated in the policy, page protection's goal is to decrease the participation of bad actors specifically while causing minimal "damage" to the unique open environment that is essential to Wikipedia. Does page protection accomplish these goals?

*8.1.1 Overview.* To begin, we overview some descriptive statistics of our treatment dataset in Table 3. On average, an article in the Wikipedia Internet Culture category is protected about 800 days after it is created, but this has a wide standard deviation (1190 days), indicating that some articles are immediately protected and others exist for years without being page protected. We also can see that page-protected articles have more activity per article than typical English Wikipedia articles in terms of edits per article (1931 vs 170) and users per article (539 vs 6.8). This is unsurprising given what we know about page protection from prior work [42] and from the policy intentions of page protection – it is intended to stop certain kinds of bad actors on articles. Bad actors likely gather on popular or contentious articles, increasing both the number and volume of expected edits.

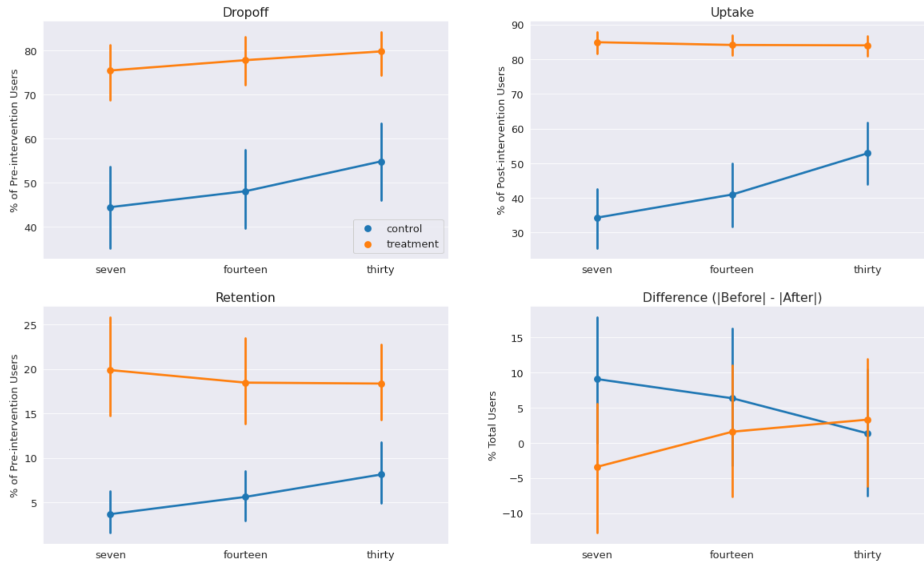


Fig. 3. Normalized participant engagement means (dropoff, uptake, retention, difference) for protected articles (orange) and comparable unprotected articles (blue). We see significantly higher amounts of dropoff and uptake in treated articles, suggesting a high amount of *user churn*. Furthermore, we see that the error bars (95% CIs) show wide dispersion in dropoff, retention, and difference across articles.

In Table 7, we present our analyses of editor engagement metrics. We first find that protected articles experience major movement in their editor landscape. These dramatic changes suggest that page protection is affecting more than the intended bad actors. In comparison to the unprotected articles in our match set, treated articles have significantly higher dropoff ($X = 2.16$, $Y = 13.84$, $p = 1.37e^{-5}$) and uptake ($X = 2.42$, $Y = 18.21$, $p = 2.35e^{-4}$).

We also consider the *relative* impact of page protection on a page, to account for any disparity in the number of editors across articles. In Figure 3, we plot our four editor engagement metrics as the percentage of eligible user populations, normalizing the effects that a given article's popularity may have on editor engagement and averaging across articles. For example, we calculate dropoff as a percentage of pre-intervention editors on a given article. For an individual article, a pre-intervention user can either be dropped or retained. Figure 3 presents the mean of the normalized user metrics (dropoff, uptake, retention, and difference) across all articles. Therefore, dropoff and retention do not account for 100% of the pre-intervention editor population, as they would on an individual article. We find that page protection impacts both pre-intervention users and post-intervention users but does so inconsistently.

*8.1.2 Consistency with Policy.* To evaluate consistency, we consider the intended goal of page protection and whether our analyses shed light on its consequences. According to Wikipedia policy, the goal of page protection is to block bad actors after a disruptive event while causing minimal damage to other editors [105]. If we represent this in desirable outcomes using our metrics, the policy intends to cause user dropoff (by removing bad actors who may come back as de-anonymized, verified users), but avoid dramatic dropoff of good-faith editors. In regards to uptake, we would expect some change, as new editors may come to a page later, but we would want to see a healthy amount of consistency in the editor landscape.

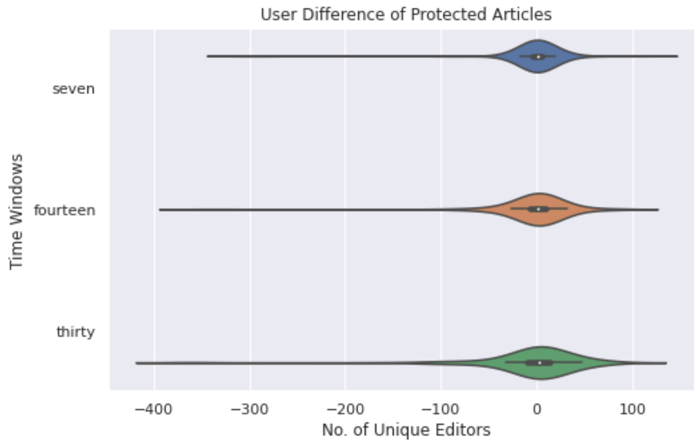


Fig. 4. Kernel density estimate of user difference amongst protected articles ($n$ = 108). The underlying distribution of user difference has an extreme spread [-400, 100], suggesting that it's unpredictable whether more or fewer users will edit a page after protection. Negative values signal that more users edited after page protection than before.

Our results suggest that protected articles experience unusually high user dropoff. As shown in both Table 5 and Figure 3, editor dropoff *increases* over time, as a typical article sees 75-85% of users who edit before the intervention do not return afterward. Moreover, these results cannot be attributed to removing IP users alone, which are a major source of bad actors on Wikipedia articles [30, 98] and a common rationale for requesting page protection (Section 6); the average percentage of anonymous users on an article is 24.75%.

These statistics show that good-faith actors meaningfully stop contributing to protected articles after protection is applied. This may be due to the controversial nature of the topic or increases in chaotic editing patterns pre-protection. However, page protection is used to calm down the editor landscape after a disruptive event. Especially thirty days after page protection, we find that dropped-off editors do not return after page protection.

Next, our results also suggest that protected articles experience significant increases in uptake, or new editors coming to the page *after* protection. Figure 3 shows that on average 85% of post-intervention editors did not contribute pre-intervention. Compared to the average uptake on comparable unprotected articles, this is a significant difference in new editors arriving on the page.

Our findings suggest that articles experience high *editor churn* after they are protected. Editor churn is where the user landscape is experiencing high turnover, with simultaneous large changes in both dropoff and uptake. For example, extreme churn is where an article has high edit activity both before and after it is protected, but there is no overlap in the set of editors pre and post-intervention. High user churn can affect community value [49] as the community members (i.e., users) are constantly changing. Specifically, on Wikipedia, a lack of repeat editors can affect the consistency of the article [77]. Additionally, our results suggest that page protection may not be meeting the stated policy goals of the intervention insofar as it (1) heavily disrupts the pre-intervention editor landscape and (2) damages the community values Wikipedia tries to uphold – which means it is likely damaging the participation of good-faith editors to the page.

*8.1.3 Consistency across articles.* Next, we examine the *consistency* of changes in editor engagement between articles. As evidenced by our control group and prior literature [78], Wikipedia articles naturally experience inconsistent editor engagement. However, these articles are all treated differently within the community. Our treatment set represents articles that received the same intervention, page protection. Moreover, the potential impact of this intervention is discussed within the community before it is instantiated. Therefore, we would expect some consistency across treatment articles to inform the appropriateness of page protection and ground deliberation. Considering our treated articles, we begin by consulting the results in Table 5, focusing on the standard deviation of the metrics we measure. Standard deviations measure the spread of data around the mean – low standard deviations indicate tight clustering around the mean, and high standard deviations indicate more spread. For page protection to be consistent, therefore, it should have a low standard deviation. Almost all of our metrics show very large standard deviations relative to the mean. For instance, 30 days after protection, the dropoff on protected articles is, on average 29.53 users, but the standard deviation is 33.31. Interpreted in a different way, 68% of protected articles had a dropoff anywhere from 0 to 62.84 (1 standard deviation from the mean), and 95% of articles have a dropoff between 0 and 129.46 (2 standard deviations from the mean). In some cases, like user difference, estimating the effects of pages that fall between 1 standard deviation away changes the sign of the effect from positive to negative. This is further demonstrated by Figure 4 which shows the dramatic spread in the underlying distribution of user difference on protected articles.

We further test this by estimating bootstrapped confidence intervals for our metrics, which estimate the population means using bootstrapped samples of the data [11]. This shows that page protection likely has inconsistent effects across articles. This is very evident for the user difference metric. Specifically, our user difference confidence interval (See Figure 3) spans negative and positive numbers ($ci_7 = [-12.7, 5.8]$) for treated articles. This means that the population mean for user difference is between -12.7 and 5.8, meaning that we cannot estimate the sign of effects. This suggests that page protection inconsistently impacts user difference, and we cannot easily predict whether there will be more editors before versus after the intervention. Interestingly, we see a similar spread in dropoff ($ci_7 = [68.91, 81.36]$) and retention ($ci_7 = [14.46, 25.7]$). In other words, when page protection is deployed we expect to see effects on both (1) the fraction of pre-intervention users who do not return after page protection and (2) the fraction of those who will continue to edit the page. Our findings suggest that it is difficult to reason about the impacts of page protection, especially when considering the impact in terms of dropoff and retention. Interestingly, uptake ($ci_7 = [81.76, 87.66]$) is the most consistent metric – we believe this shows that page protection leads to new editors moving to the page.

In summary, our results in this section suggest that effects on page protection are mostly inconsistent for pre-intervention editors but not those who entered the article after. We further validate these results by performing a robustness check on articles in the Ongoing Conflicts category on Wikipedia (Appendix Figure 6) and find that those articles similarly experience unusually high and inconsistent editor movement.

### 8.2 RQ3. How does page protection impact contributor concentration?

Next, we move to RQ3, which addresses editor concentration. A primary purpose of Wikipedia is to leverage the "wisdom of the crowd" to generate and refine content to meet its high editorial standards [7]. While Wikipedia's policies clearly value open contribution [61, 116], prior work has also highlighted how the crowd's contributions are heavily dominated by a select few power users [56, 77] which contributes to Wikipedia's intimidating hierarchy [14, 82] and "hidden order" [98].

Page protection is a redistributive intervention: it changes the opportunity to edit a page amongst the editor population by preventing some editors from participating (most often, anonymous users

| | | Hoover | | 20:20 | |
|---|---|---|---|---|---|
| | Conditions | Mean(SD) | p-val | Mean(SD) | p-val |
| **match** | pre | 0.34(0.15) | – | 7.32(6.98) | – |
| | seven | 0.31(0.41) | 0.359 | 7.35(7.04) | 0.724 |
| | fourteen | 0.36(0.42) | 0.628 | 7.68(7.28) | **0.012**$^{*}$ |
| | thirty | 0.43(0.41) | **0.037**$^{*}$ | 7.97(7.01) | **0.008**$^{*}$ |
| **protected** | pre | 0.38(0.12) | – | 4.99(3.01) | – |
| | seven | 0.84(0.16) | **1.31E-46**$^{***}$ | 5.43(2.74) | **.014**$^{*}$ |
| | fourteen | 0.81(0.17) | **1.91E-41**$^{***}$ | 6.09(3.38) | **3.82E-05**$^{***}$ |
| | thirty | 0.771(0.18) | **1.11E-35**$^{***}$ | 6.74(3.8) | **4.01E-08**$^{***}$ |

Table 8. $p$-val indicates statistical significance between the pre-protection edit shares and the $n$ days post-protection edit shares. We find $p < .05$ for all protected conditions. The average Hoover indices dramatically increase in the treatment set *seven* days after page protection. Significance codes: p-value <0.001 '$***$', 0.001 '$**$', 0.05 '$*$'

and those with few accepted edits). However, page protection's goal is to provide the minimum effective dose of redistribution – its goal is to minimize the impact on good-faith editors. Therefore, when page protection is enacted, we would expect to see some increases in measures of concentration (because by its nature, protection limits participation).

We analyze two inequality metrics with respect to the share of total edits per contributor on protected articles: the Hoover index and the 20:20 score. Recall from the Methods (and Table 6) that the Hoover index gives insights into the *amount* of inequality while the 20:20 ratio helps us understand *where* in the population that disparity comes from.

| | Hoover | 20:20 |
|---|---|---|
| **pre** | 0.034$^{*}$ | 0.002$^{*}$ |
| **seven** | 2.78E-26$^{***}$ | 0.009$^{*}$ |
| **fourteen** | 3.19E-19$^{***}$ | 0.042$^{*}$ |
| **thirty** | 2.61E-13$^{***}$ | 0.111 |

Table 9. Presents significance tests of means ($t$-tests) between control and treatment sets. All conditions show statistical significance except for the 20:20 ratio thirty days after page protection. This suggests that page protection increases overall inequality, but doesn't increase disparity between heavy and light editors. Significance codes: p-value <0.001 '$***$', 0.001 '$**$', 0.05 '$*$'

We first explore how inequality manifests on protected pages, especially in relation to comparable unprotected pages (our match set). In Table 9, we show the statistical tests of the difference of means between protected articles and our match set. Protected articles in the Internet Culture category have significantly different inequality across both parity metrics (Hoover: $X = .31, Y = .84, p = 1.78e^{-26}$; 20:20: $X = 7.35, Y = 5.34, p = .009$). We do note that page protected articles have more unequal editor parity *before* page protection, as measured by the Hoover index ($X = .34, Y = .38, p = .034$). This is reasonable to expect, even when pre-protection metrics (pageviews, edits, etc) are controlled for, given that protecting a page is reactive to the condition of a given page. According to the

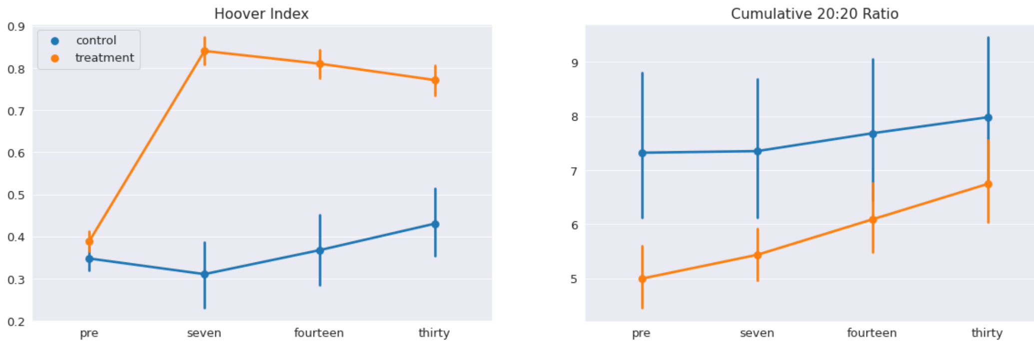


Fig. 5. Contributor concentration trends (Hoover Index and Cumulative 20:20 Ratio) for protected articles (orange) and matched articles (blue). We see increases in all three metrics amongst treated articles. Specifically, we see dramatic increases in the Hoover index. Error bars indicate 95% CIs.

policy page, "Pages are protected when a specific damaging event has been identified that cannot be prevented through other means." [105] Said another way, before a page is protected, there is already something abnormal happening, whether that is controversial topics or contention between editors.

Next, we consider how editor concentration on protected articles compares to our match set of unprotected ones. We present the inequality measures and their comparisons to both unprotected data and the three time windows in our analysis in Table 8. Recall that we expect to see *some increase* in inequality amongst contributors after page protection to indicate that bad actors are no longer allowed to contribute. To examine this, we use the Hoover Index. The Hoover index represents how many resources would need to be redistributed to achieve perfect equality and, therefore, suggests how powerful a perfectly-equal solution would have to be.

Our results suggest that protected pages experience significantly larger amounts of inequality after the intervention. Within our set of treated articles, we see the Hoover index significantly increase even within just seven days of page protection ($X = .38$, $Y = .84$, $p = 1.31e^{-46}$). This implies that page protection as implemented on Wikipedia likely has immediate effects on contributor concentration. This suggests that page protection concentrates editing to a small number of editors.

Putting these measures into the context of reducing inequality, we see that achieving equality within these pages would require drastic redistribution. For instance, Table 8 shows that seven days after page protection 84% of edits would need to be redistributed amongst the editor population to achieve perfect equality. While a completely equal contribution landscape is not achievable, this is a dramatic contrast from our unprotected match set where only 28.9% of edits would have to be redistributed.

Finally, we consider where in the editor population is this disparity coming from. We do so by analyzing the cumulative 20:20 ratio, which is the ratio of contributions amongst the top 20% of editors to the bottom 20%. In the context of a Wikipedia article, this helps us measure the disparity between heavy, power users and light users, giving insights into whether Wikipedia contributions are motivated by the "power of the few" [56]. Figure 5 shows that the 20:20 ratio of protected articles increases over time. However, over time, the 20:20 ratio among treated articles is the same level as with unprotected articles ($X = 7.97$, $Y = 6.74$, $p = .111$). In other words, before page protection, we see that heavy editors are editing nearly 5 times as much as light editors. Thirty days after page protection, this increases to nearly 7 times, but is on par with unprotected articles. Given that this is a cumulative score (i.e., edit shares are compounded over time), our results suggest that the top-20% of editors pre-protection continue dominating the editing landscape. However, we eventually see no significant difference in 20:20 ratios across experimental conditions, suggesting that page protection does not cause the "rich to get richer" in the mid-term.

To summarize, when evaluating editor concentration on page protected articles, we see mixed outcomes. Our results suggest that page protection does not amplify the disparity between heavy and light editors. This is especially desirable for protected articles because they are oftentimes more controversial or higher-profile articles. At the same time, Hoover indices still demonstrate a huge spike in contribution inequality. While the 20:20 ratio demonstrates that the power users do not get substantively more powerful, there is still a dramatic amount of editors that lose their power to edit as evidenced by the scope of the Hoover measure. Our robustness check on articles in the Ongoing Conflicts category mirrors this dramatic increase in the Hoover index with little change to the 20:20 ratio (Appendix Figure 7).

## 9 DISCUSSION

Our research has explored how the editor landscape changes with respect to page protection on Wikipedia. Specifically, we look at two dimensions that are heavily correlated to Wikipedia's core values: (1) participant engagement and (2) contributor concentration. We find that the consequences of page protection run counter to Wikipedia's values and policies. In RQ2, we found that protected pages experience substantial editor churn, implying that the editor population almost completely turns over. These outcomes are not consistent nor predictable for a page. In RQ3, we find that protected pages experience increased inequity in contributor concentration. In this section, we discuss how our findings contextualize current conversations surrounding frictions, moderator interventions, and peer participation.

### 9.1 Governance and New Dimensions for Measuring Friction

HCI has a rich history of exploring design frictions [22, 75, 102]. Historical work has long encouraged creating "seamless interactions" to promote positive user experiences [60, 65] – making systems and interactions easier, faster, and more effective to use. On Wikipedia especially, previous literature has heavily focused on making contributions easier overall with an explicit focus on newcomers [20, 40, 73] as well as experts [101, 117, 121].

However, other research has challenged this need for seamlessness and has considered how the introduction of strategic frictions can serve design goals and values. On Wikipedia, frictions are often used to create and uphold community values. While we explore page protection specifically, edit reversion [40], participant bans [30], and citation standards [6] are mechanisms that make it harder to contribute to Wikipedia, but help maintain the platform's integrity.

Considering friction first from its "theoretical" or policy perspective, our findings suggest that page protection in the Internet Culture category is a friction with mixed effects that are not consistent with the platform's desired consequences. The Wikipedia page protection policy suggests that the intervention is meant to block contributions from bad actors with minimal damage to good faith editors [105]. However, our results in RQ2 show that protected pages experience high editor churn, resulting in a new wave of editors arriving on a page after it is protected. Moreover, we show that page protection is an *inconsistent* friction, with its specific effect being hard to precisely estimate. In essence, page protection does dramatically change the editor landscape on a given page on average, but it is hard to reason or predict what protection will do on a given page.

Our work here joins other work on how the policy intentions of frictions may not have the intended effects. For example, Chancellor et al. [16] found that content moderation frictions on sharing dangerous eating disorder content did not slow the community in participation, and in fact may have made it worse. Likewise, we have seen that user habituation to pop-ups leads to ineffective frictions where people ignore the pop-up [2, 24]. Even Wikipedia research acknowledges that there is “friction” in becoming a Wikipedian [13, 77].

These results suggest that friction cannot be simply treated as a binary of its effect. Many frictions, such as community bans are thought of as on/off – user banning, removing content, reverting edits, or blocking participation. To truly capture the effects of friction, we argue that researchers and designers expand our perspective on what friction is, how it operates, as well as new ways to measure it. Our results highlight how the same intervention can have a massively different effect on different articles. What if, rather than conceptualize a single intervention the same way, we take inspiration from physics to evaluate an intervention’s *coefficient of friction* on a given article? Friction lies on a spectrum in terms of how many users it impacts and to what extent it affects them. This is important in platform governance because it can be difficult to build systematic policies around interventions when the effects may be inconsistent with the intervention’s intention or inconsistent across subjects.

The Wikipedia page protection policy already alludes to this concept by documenting that page protection should block bad actors while causing minimal damage to others. In other words, page protection should have a high coefficient of friction on bad actors, giving them opportunities to verify themselves and continue contributing, but a low coefficient on others. Placing the policy in these quantitative terms of high and low friction could better assist in measuring differential effects, and inform discussions about whether the policy goals are being achieved. Likewise, coefficients could be a useful way for community managers and administrators to better conceptualize the impacts of friction ahead of time and monitor a policy’s impacts, allowing community leaders to be more nimble in assessing the outcome of a given intervention.

Our work enables platform policymakers to associate quantifiable metrics with their intervention goals. Specifically, placing goals in terms of *how much* friction should be instantiated and *who* should feel the strongest effects.

### 9.2 Bluntness and Measuring Precision of Broad Interventions

Recent work on content moderation has focused on studying the effects of broad moderator interventions, such as subreddit banning [18], deplatforming [47], and hashtag banning [16]. From a policy perspective, page protection shares traits with sitewide interventions insofar as it (1) intends to block bad actors and (2) is targeted at a broad swathe of users rather than any particular individual. Note that with page protection, blocked users have the opportunity to continue contributing if they verify themselves on the platform. We understand and agree with much of the intentions behind the page protection policy on Wikipedia: it is valuable to pause, stop, or freeze user contributions when these contributions damage the quality of something as important as an encyclopedia article. In this way, friction acts as a proverbial “emergency break.”

However, we argue that page protection is a *blunt instrument*, in that it is both heavy-handed and imprecise in its effects. The lowest level of edit protection blocks IP users from editing a specific article [42, 105]. Our results suggest that articles with this protection have an excessive number of users to drop off that exceeds the number of IP users. Moreover, we see a high amount of user churn in articles with page protection. This suggests that page protection is affecting more than just the intended editors – meaning it is not nuanced in how it captures behavior. In fact, page protection blocks people based on the categories they fall into, not based on their past behaviors. This runs counter to many other Wikipedia-enforced frictions on the site, which focus on banning people based on behavior (like vandalism [30] and potential damage of an edit with ORES [38]). Our findings question the effectiveness of blunt interventions like page protection in enacting the stated goals of the platform.

How could platforms like Wikipedia and other peer production platforms limit the bluntness of moderator strategies through design interventions? Many peer production platforms use precise interventions, such as the removal of a single post or comment [47], in tandem with more blunt interventions, such as community banning [17]. One design implication of our work is that platform administrators should consider adding safeguards to ensure blunt interventions do not have undesirable adverse effects, or that they can be corrected. Similarly, more transparency and explanation of decision-making could help explain why a policy is being implemented and, in the long run, inform discussion and encourage policy change. In the case of Reddit, feelings of opaqueness can lead to disruptive discussion amongst community members [48]. Vaccaro et al. [97] discusses the importance of contestability in content moderation for more participatory communities. For Wikipedia and page protection, contestability of decisions and protection could ensure that more people are brought into deliberation about decisions. Although Wikipedia already has mechanisms for post-intervention discussions, such as talk pages, our work underpins the need for wider adoption of deliberative venues on self-governed platforms.

### 9.3 Peers Are the Producers of Frictions: Future Work on Participatory Decisions of Friction

Community-based governance is a core value to Wikipedia [70, 85] and other peer production platforms [8, 80]. In particular, Wikipedia relies on peers for both content production and governance [86]. However, because peers run the platform, this also means that "peers" produce the frictions that happen on Wikipedia, including page protection [50]. Peers on Wikipedia deliberate on many facets of the site, including deletion [67], creation [10], and disputing edit quality [74]. This is in contrast to many social platforms which are studied in HCI and CSCW, where enforcement of friction ultimately is done by a corporate platform.

However, peer production platforms do exert non-democratic means of control on their platforms. Prior work on Wikipedia has shown that, in attempting to manage an open and growing platform, Wikipedia has inadvertently created strict hierarchies [14, 82, 88], a "hidden order" of tacit rules [98], and inequality amongst contributors [76]. Our work explores how these ideas manifest after a governance intervention is instantiated. Our results indicate that page protection increases inequality in contributor concentration amongst pages that were protected. We showed that page protection may lead to editor inequity and, therefore, may exacerbate some of the outcomes that Wikipedia is actively trying to combat.

Our findings raise complex questions about the relevant stakeholders in participatory interventions; page protection includes Wikipedia editors and administrators but not the subjects of the articles. Historically, excluding certain stakeholders through page protection has been controversial for Wikipedia. Gamergate (2014) was a harassment campaign against the intersection of feminism and video game culture. It is notably cited as one of the best-documented incidents of large-scale bullying behavior online [66] and lead to Wikipedia taking blocking actions–such as page protection, editor sanctioning, and editor banning–on relevant articles. At the height of the controversy, Mark Bernstein, former Wikipedia editor, said that Wikipedia's decision places the power into Gamergaters' hands, "not only do the Gamergaters get to rewrite their own page (and Zoe Quinn's, Brianna Wu's, Anita Sarkeesian's, etc); feminists are to be purged en bloc from the encyclopedia." [41] Gamergate highlights how those affected by the article's content may also exist as Wikipedia editors.

When we put this history in the context of our results that page protection affects editor concentration, we highlight the value of participatory governance mechanisms, such as deliberative juries [25] or other community-based interventions [86]. We build off these calls for participatory content moderation by suggesting that we expand who we choose to participate. Specifically thinking about content surrounding Internet Culture, where it would be feasible to involve subjects, we propose that researchers consider involving article subjects as participants.

### 9.4 Limitations and Future Work

One limitation of our study is that we explored a single category on Wikipedia: Internet Culture. This was due in part to topical relevance as well as the difficulty of detecting protection events on articles that are not currently protected (see similar challenges by Hill and Shaw [42]). The Wikipedia Internet Culture category is uniquely participatory and has been a venue of controversy, such as the Gamergate article dispute. Given our robustness check (see Appendix), we are confident that our findings hold for more than just our category, but we cannot be sure of this. Future work could look at how additional topic dimensions–such as technicality, relevance to pop culture, etc.–interact with page protection and similar kinds of interventions. Furthermore, we took a policy-based approach to understand the platform's model of the intervention. However, we do not explore how Wikipedia administrators perceive page protection or privately deliberate about its effectiveness. Another limitation of our approach is that our work is observational, and is limited because of the difficulty in constructing a true quasi-causal analysis with more causal outcomes (see Methods). Although there is ample evidence of observational work on frictions in CSCW [16] and HCI [18, 47], the lack of truly causal methods limits our claims. If more data about the page protection evaluation process becomes available in the future (e.g., via new data releases or because the process changes), it may be possible to build on our and directly estimate causal treatment effects using RDD, ITS, or other methods.

Additionally, our work focuses solely on the consequences of page protection to the editor landscape. To better understand the tradeoffs involved with protected an article, future work could examine the effectiveness of page protection. Using the quasi-causal methods mentioned above, researchers could understand how page protection impacts article quality.

Future work could interview Wikipedia stakeholders to further understand page protection, why it occurs, and how editors feel about being blocked by this friction. Future research could also compare protection to other frictions, operationalizing the idea of "friction coefficients." Finally, we also envision participatory system design [38] to create tools that may make protection more consistent or in line with policy.

## 10 CONCLUSION

In this paper, we offered a mixed methods analysis of the motivations and consequences of page protection in Wikipedia's Internet Culture category. Overall, we observed a misalignment in the Wikipedia policies versus the consequential effects. While this protection policy highlights the need for minimal damage to the editor landscape, our results suggest that page protection has more complex effects than simply accomplishing a policy goal or not. Our analysis showed that page protection's effects on editor engagement are inconsistent across articles, increasing user churn, and, therefore, hard to consistently estimate. Furthermore, page protection can have unintended effects by increasing the contributor concentration amongst the editor population. Our findings raise interesting questions as to how platforms should negotiate the tension between promoting participation and limiting poor citizenship, and we hope that peer production platforms consider our ideas in helping them more effectively implement policies that impact their contributor populations.

## ACKNOWLEDGMENTS

This work is derived from observations of publicly-available Wikipedia data and is indebted to the work of that community.

## REFERENCES


[1] 2017. Facebook is ditching its own solution to fake news because it didn't work. https://qz.com/1162973/to-fight-fake-news-facebook-is-replacing-flagging-posts-as-disputed-with-related-articles/. Accessed: 2022-8-25.

[2] Ammar Amran, Zarul Fitri Zaaba, and Manmeet Kaur Mahinderjit Singh. 2018. Habituation effects in computer security warning. *Information Security Journal: A Global Perspective* 27, 4 (July 2018), 192–204.

[3] Peter C Austin. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research* 46, 3 (2011), 399–424.

[4] Martina Balestra, Coye Cheshire, Ofer Arazy, and Oded Nov. 2017. Investigating the Motivational Paths of Peer Production Newcomers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6381–6385.

[5] Amanda Baughan, Mingrui Ray Zhang, Raveena Rao, Kai Lukoff, Anastasia Schaadhardt, Lisa D Butler, and Alexis Hiniker. 2022. "I Don't Even Remember What I Read": How Design Influences Dissociation on Social Media.

[6] Omer Benjakob, Rona Aviram, and Jonathan Aryeh Sobel. 2022. Citation needed? Wikipedia bibliometrics during the first wave of the COVID-19 pandemic. *Gigascience* 11, 1 (Jan. 2022).

[7] Benkler, Shaw, and Hill. 2015. Peer production: A form of collective intelligence. In *Handbook of Collective Intelligence*, Thomas Malone and Michael Bernstein (Eds.). MIT Press.

[8] Yochai Benkler. 2002. Coase's penguin, or, linux and" the nature of the firm". *Yale law journal* (2002), 369–446.

[9] Yochai Benkler and Helen Nissenbaum. 2006. Commons-based peer production and virtue. *Journal of political philosophy* 14, 4 (2006).

[10] Taryn Bipat, Diana Victoria Davidson, Melissa Guadarrama, Nancy Li, Ryder Black, David W McDonald, and Mark Zachry. 2019. How does Editor Interaction Help Build the Spanish Wikipedia?. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 156–160.

[11] K D Bird. 2004. *Analysis of Variance via Confidence Intervals.* SAGE.

[12] Joshua E Blumenstock. 2008. Size matters: word count as a measure of quality on wikipedia. In *Proceedings of the 17th international conference on World Wide Web* (Beijing, China) *(WWW '08)*. Association for Computing Machinery, New York, NY, USA, 1095–1096.

[13] Susan L Bryant, Andrea Forte, and Amy Bruckman. 2005. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*. 1–10.

[14] Brian Butler, Elisabeth Joyce, and Jacqueline Pike. 2008. Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1101–1110.

[15] Stevie Chancellor and Scott Counts. 2018. Measuring employment demand using internet search data. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[16] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. #thyghgapp. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco California USA). ACM, New York, NY, USA.

[17] Eshwar Chandrasekharan, Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2022. Quarantined! Examining the Effects of a Community-Wide Moderation Intervention on Reddit. *ACM Trans. Comput.-Hum. Interact.* 29, 4 (March 2022), 1–26.

[18] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (Dec. 2017), 1–22.

[19] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. Antisocial Behavior in Online Discussion Communities. *ICWSM* 9, 1 (2015), 61–70.

[20] Giovanni Luca Ciampaglia and Dario Taraborelli. 2015. MoodBar: Increasing New User Retention in Wikipedia through Lightweight Socialization. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 734–742.

[21] Victoria Clarke, Virginia Braun, and Nikki Hayfield. 2015. Thematic analysis. *Qualitative psychology: A practical guide to research methods* 222, 2015 (2015), 248.

[22] Anna L Cox, Sandy J J Gould, Marta E Cecchinato, Ioanna Iacovides, and Ian Renfree. 2016. Design Frictions for Mindful Interactions: The Case for Microboundaries. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1389–1397.

[23] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2098–2110.

[24] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. 2008. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1065–1074.

[25] Jenny Fan and Amy X Zhang. 2020. Digital Juries: A Civics-Oriented Approach to Platform Governance.

[26] Casey Fiesler, Shannon Morrison, and Amy S Bruckman. 2016. An Archive of Their Own: A Case Study of Feminist HCI and Values in Design. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2574–2585.

[27] Heather Ford and Judy Wajcman. 2017. 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. *Soc. Stud. Sci.* 47, 4 (Aug. 2017), 511–527.

[28] Vijaya Gadde and Kayvon Beykpour. 2020. Additional steps we're taking ahead of the 2020 US Election. https://blog.twitter.com/en_us/topics/company/2020/2020-election-changes.

[29] R Stuart Geiger and Aaron Halfaker. 2013. When the levee breaks: without bots, what happens to Wikipedia's quality control processes?. In *Proceedings of the 9th International Symposium on Open Collaboration* (Hong Kong, China) *(WikiSym '13, Article 6)*. Association for Computing Machinery, New York, NY, USA, 1–6.

[30] R Stuart Geiger and David Ribes. 2010. The work of sustaining order in wikipedia: the banning of a vandal. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (Savannah, Georgia, USA) *(CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 117–126.

[31] Sarah A Gilbert. 2020. "I run the world's largest historical outreach project and it's on a cesspool of a website." Moderating a Public Scholarship Site on Reddit: A Case Study of r/AskHistorians. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1 (May 2020), 1–27.

[32] S J J Gould, L L Chuang, I Iacovides, and others. 2021. A Special Interest Group on Designed and Engineered Friction in Interaction. *Extended Abstracts of* (2021).

[33] Colin M Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18, Paper 534)*. Association for Computing Machinery, New York, NY, USA, 1–14.

[34] James Grimmelmann. 2017. The virtues of moderation. (May 2017).

[35] Barbara Grosse-Hering, Jon Mason, Dzmitry Aliakseyeu, Conny Bakker, and Pieter Desmet. 2013. Slow design for meaningful interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3431–3440.

[36] Barbara Grosse-Hering, Jon Mason, Dzmitry Aliakseyeu, Conny Bakker, and Pieter Desmet. 2013. Slow design for meaningful interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 3431–3440.

[37] Hussam Habib, Maaz Bin Musa, Muhammad Fareed Zaffar, and Rishab Nithyanand. 2022. Are Proactive Interventions for Reddit Communities Feasible? *ICWSM* 16 (May 2022), 264–274.

[38] Aaron Halfaker and R Stuart Geiger. 2020. ORES: Lowering Barriers with Participatory Machine Learning in Wikipedia. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2 (Oct. 2020), 1–37.

[39] Aaron Halfaker, R Stuart Geiger, Jonathan T Morgan, and John Riedl. 2013. The rise and decline of an open collaboration system. *Am. Behav. Sci.* 57, 5 (May 2013), 664–688.

[40] Aaron Halfaker, Aniket Kittur, and John Riedl. 2011. Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (Mountain View, California) *(WikiSym '11)*. Association for Computing Machinery, New York, NY, USA, 163–172.

[41] Alex Hern. 2015. Wikipedia votes to ban some editors from gender-related articles. *The Guardian* (Jan. 2015).

[42] Benjamin Mako Hill and Aaron Shaw. 2015. Page protection: another missing dimension of Wikipedia research. In *Proceedings of the 11th International Symposium on Open Collaboration* (San Francisco, California) *(OpenSym '15, Article 15)*. Association for Computing Machinery, New York, NY, USA, 1–4.

[43] Benjamin Mako Hill and Aaron Shaw. 2021. The Hidden Costs of Requiring Accounts: Quasi-Experimental Evidence From Peer Production. *Communic. Res.* 48, 6 (Aug. 2021), 771–795.

[44] Mo Houtti, Isaac Johnson, Joel Cepeda, Soumya Khandelwal, Aviral Bhatnagar, and Loren Terveen. 2022. " We Need a Woman in Music": Exploring Wikipedia's Values on Article Priority. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–28.

[45] Stefano M Iacus, Gary King, and Giuseppe Porro. 2012. Causal Inference without Balance Checking: Coarsened Exact Matching. *Polit. Anal.* 20, 1 (2012), 1–24.

[46] Farnaz Jahanbakhsh, Amy X Zhang, Adam J Berinsky, Gordon Pennycook, David G Rand, and David R Karger. 2021. Exploring Lightweight Interventions at Posting Time to Reduce the Sharing of Misinformation on Social Media. *Proc.*

*ACM Hum.-Comput. Interact.* 5, CSCW1 (April 2021), 1–42.
[47] Shagun Jhaver, Christian Boylston, Diyi Yang, and Amy Bruckman. 2021. Evaluating the Effectiveness of Deplatforming as a Moderation Strategy on Twitter. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 1–30.
[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 Hum.-Comput. Interact.* 3, CSCW (Nov. 2019), 1–27.
[49] Marcel Karnstedt, Matthew Rowe, Jeffrey Chan, Harith Alani, and Conor Hayes. 2011. The effect of user features on churn in social networks. In *Proceedings of the 3rd International Web Science Conference* (Koblenz, Germany) *(WebSci '11, Article 23)*. Association for Computing Machinery, New York, NY, USA, 1–8.
[50] Brian Keegan and Casey Fiesler. 2017. The evolution and consequences of peer producing wikipedia's rules. In *Eleventh International AAAI Conference on Web and Social Media*.
[51] Brian C Keegan. 2013. A history of newswork on Wikipedia. In *Proceedings of the 9th International Symposium on Open Collaboration* (Hong Kong, China) *(WikiSym '13, Article 7)*. Association for Computing Machinery, New York, NY, USA, 1–10.
[52] Brian C Keegan and Jed R Brubaker. 2015. 'is' to 'was'. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver BC Canada). ACM, New York, NY, USA.
[53] Brian C Keegan and Chenhao Tan. 2020. A Quantitative Portrait of Wikipedia's High-Tempo Collaborations during the 2020 Coronavirus Pandemic. (June 2020). arXiv:2006.08899 [cs.SI]
[54] Johannes Kiesel, Martin Potthast, Matthias Hagen, and Benno Stein. 2017. Spatio-Temporal Analysis of Reverted Wikipedia Edits. *ICWSM* 11, 1 (May 2017), 122–131.
[55] Yongnam Kim and Peter Steiner. 2016. Quasi-Experimental Designs for Causal Inference. *Educ. Psychol.* 51, 3-4 (Sept. 2016), 395–405.
[56] Kittur, Chi, Pendleton, Suh, and Mytkowicz. 2007. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World Wide Web J. Biol.* (2007).
[57] Kollin. 2018. Designing friction for a better user experience. *Smashing Magazine* (2018).
[58] Srijan Kumar, Francesca Spezzano, and V S Subrahmanian. 2015. VEWS: A Wikipedia Vandal Early Warning System. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Sydney, NSW, Australia) *(KDD '15)*. Association for Computing Machinery, New York, NY, USA, 607–616.
[59] Shyong (tony) K Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R Musicant, Loren Terveen, and John Riedl. 2011. WP:clubhouse? an exploration of Wikipedia's gender imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (Mountain View, California) *(WikiSym '11)*. Association for Computing Machinery, New York, NY, USA, 1–10.
[60] Jonathan Lazar, Adam Jones, and Ben Shneiderman. 2006. Workplace user frustration with computers: an exploratory investigation of the causes and severity. *Behav. Inf. Technol.* 25, 3 (May 2006), 239–251.
[61] Bo Leuf and Ward Cunningham. 2001. *The Wiki way: quick collaboration on the Web.* Addison-Wesley Longman Publishing Co., Inc., USA.
[62] H Li, B Hecht, and S Chancellor. 2022. All That's Happening behind the Scenes: Putting the Spotlight on Volunteer Moderator Labor in Reddit. *of the International AAAI Conference on ...* (2022).
[63] Larry Long and Alfred Nucci. 1997. The Hoover Index of Population Concentration: A Correction and Update. *Prof. Geogr.* 49, 4 (Nov. 1997), 431–440.
[64] Valerio Lorini, Javier Rando, Diego Saez-Trumper, and Carlos Castillo. 2020. Uneven Coverage of Natural Disasters in Wikipedia: the Case of Flood. (Jan. 2020). arXiv:2001.08810 [cs.IR]
[65] Martin Maguire. 2014. Socio-technical systems and interaction design - 21st century relevance. *Appl. Ergon.* 45, 2 (March 2014), 162–170.
[66] Adrienne Massanari. 2017. #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society* 19, 3 (March 2017), 329–346.
[67] Elijah Mayfield and Alan W Black. 2019. Analyzing Wikipedia Deletion Debates with a Group Decision-Making Forecast Model. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.
[68] Brian McInnis, Leah Ajmani, Lu Sun, Yiwen Hou, Ziwen Zeng, and Steven P Dow. 2021. Reporting the Community Beat: Practices for Moderating Online Discussion at a News Website. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 1–25.
[69] Amanda Menking, Ingrid Erickson, and Wanda Pratt. 2019. People Who Can Take It: How Women Wikipedians Negotiate and Navigate Safety. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19, Paper 472)*. Association for Computing Machinery, New York, NY, USA, 1–14.
[70] Amanda Menking and Jon Rosenberg. 2021. WP:NOT, WP:NPOV, and other stories Wikipedia tells us: A feminist critique of Wikipedia's epistemology. *Sci. Technol. Human Values* 46, 3 (May 2021), 455–479.

[71] David Meyer. 2021. Facebook whistleblower Frances Haugen tells lawmakers the only way to fix the company is to partially destroy its business model. https://fortune.com/2021/10/25/facebook-business-model-selective-friction-frances-haugen-whistleblower-uk-parliament/. Accessed: 2022-8-30.
[72] Marc Miquel-Ribé and David Laniado. 2020. The Wikipedia Diversity Observatory: A Project to Identify and Bridge Content Gaps in Wikipedia. In *Proceedings of the 16th International Symposium on Open Collaboration* (Virtual conference, Spain) *(OpenSym 2020, Article 14)*. Association for Computing Machinery, New York, NY, USA, 1–4.
[73] Jonathan T Morgan and Aaron Halfaker. 2018. Evaluating the impact of the Wikipedia Teahouse on newcomer socialization and retention. In *Proceedings of the 14th International Symposium on Open Collaboration* (Paris, France) *(OpenSym '18, Article 20)*. Association for Computing Machinery, New York, NY, USA, 1–7.
[74] Sneha Narayan, Nathan TeBlunthuis, Wm Salt Hale, Benjamin Mako Hill, and Aaron Shaw. 2019. All Talk: How Increasing Interpersonal Communication on Wikis May Not Enhance Productivity. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (Nov. 2019), 1–19.
[75] Don Norman. 2013. *The Design of Everyday Things: Revised and Expanded Edition.* Basic Books.
[76] Felipe Ortega, Jesus M Gonzalez-Barahona, and Gregorio Robles. 2008. On the Inequality of Contributions to Wikipedia. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*. 304–304.
[77] Katherine Panciera, Aaron Halfaker, and Loren Terveen. 2009. Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work*. 51–60.
[78] Katherine Panciera, Aaron Halfaker, and Loren Terveen. 2009. Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work* (Sanibel Island, Florida, USA) *(GROUP '09)*. Association for Computing Machinery, New York, NY, USA, 51–60.
[79] Hao Peng, Daniel M Romero, and Emőke-Ágnes Horvát. 2022. Dynamics of cross-platform attention to retracted papers. *Proc. Natl. Acad. Sci. U. S. A.* 119, 25 (June 2022), e2119086119.
[80] Eric Raymond. 1999. S. The cathedral and the bazaar: musings on Linux and Open Source by an accidental revolutionary.
[81] Miriam Redi, Besnik Fetahu, Jonathan Morgan, and Dario Taraborelli. 2019. Citation Needed: A Taxonomy and Algorithmic Assessment of Wikipedia's Verifiability. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1567–1578.
[82] Emiel Rijshouwer, Justus Uitermark, and Willem de Koster. 2021. Wikipedia: a self-organizing bureaucracy. *Inf. Commun. Soc.* (Oct. 2021), 1–18.
[83] Adi Robertson. 2015. Wikipedia denies 'purging' feminist editors over Gamergate debate. https://www.theverge.com/2015/1/28/7927425/wikipedia-bans-gamergate-editors-violating-policies. Accessed: 2022-12-6.
[84] Diego Saez-Trumper. 2019. Online Disinformation and the Role of Wikipedia. (Oct. 2019). arXiv:1910.12596 [cs.CY]
[85] Seering. 2020. Reconsidering community self-moderation: the role of research in supporting community-based models for online content moderation. *Proc. ACM Hum. Comput. Interact.* (2020).
[86] Joseph Seering. 2020. Reconsidering Self-Moderation: the Role of Research in Supporting Community-Based Models for Online Content Moderation. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2 (Oct. 2020), 1–28.
[87] Joseph Seering, Geoff Kaufman, and Stevie Chancellor. 2022. Metaphors in moderation. *New Media & Society* 24, 3 (March 2022), 621–640.
[88] Aaron Shaw and Benjamin M Hill. 2014. Laboratories of oligarchy? How the iron law extends to peer production. *Journal of Communication* 64, 2 (2014), 215–238.
[89] Strauss and Fuad-Luke. 2008. The slow design principles. *Proceedings of the Changing the Change* (2008).
[90] Elizabeth A Stuart. 2010. Matching methods for causal inference: A review and a look forward. *Stat. Sci.* 25, 1 (Feb. 2010), 1–21.
[91] Kelsey Suyehira and Francesca Spezzano. 2016. DePP: A System for Detecting Pages to Protect in Wikipedia. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* (Indianapolis, Indiana, USA) *(CIKM '16)*. Association for Computing Machinery, New York, NY, USA, 2081–2084.
[92] Nathan TeBlunthuis, Aaron Shaw, and Benjamin Mako Hill. 2018. Revisiting "The Rise and Decline" in a Population of Peer Production Projects. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18, Paper 355)*. Association for Computing Machinery, New York, NY, USA, 1–7.
[93] Misha Teplitskiy, Grace Lu, and Eamon Duede. 2017. Amplifying the impact of open access: Wikipedia and the diffusion of science. *J. Assoc. Inf. Sci. Technol.* 68, 9 (Sept. 2017), 2116–2127.
[94] Ija Trapeznikova and Royal Holloway University of London, UK. 2019. Measuring income inequality. *IZA World Labor* (2019).
[95] Amaury Trujillo and Stefano Cresci. 2022. Make Reddit Great Again: Assessing Community Effects of Moderation Interventions on r/The_Donald. (Jan. 2022). arXiv:2201.06455 [cs.SI]
[96] Milena Tsvetkova, Ruth García-Gavilanes, Luciano Floridi, and Taha Yasseri. 2017. Even good bots fight: The case of Wikipedia. *PLoS One* 12, 2 (Feb. 2017), e0171774.

[97] Kristen Vaccaro, Ziang Xiao, Kevin Hamilton, and Karrie Karahalios. 2021. Contestability For Content Moderation. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2 (Oct. 2021), 1–28.
[98] Fernanda B Viégas, Martin Wattenberg, and Matthew M McKeon. 2007. The Hidden Order of Wikipedia. In *Online Communities and Social Computing*. Springer Berlin Heidelberg, 445–454.
[99] N Vincent and B Hecht. 2021. A deeper investigation of the importance of wikipedia links to search engine results. *Proceedings of the ACM on Human-Computer* (2021).
[100] William Yang Wang and Kathleen R McKeown. 2010. "Got you!": automatic vandalism detection in Wikipedia with web-based shallow syntactic-semantic modeling. In *Proceedings of the 23rd International Conference on Computational Linguistics* (Beijing, China) *(COLING '10)*. Association for Computational Linguistics, USA, 1146–1154.
[101] Shira Schecter Weiner, Jill Horbacewicz, Lane Rasberry, and Yocheved Bensinger-Brody. 2019. Improving the Quality of Consumer Health Information on Wikipedia: Case Series. *J. Med. Internet Res.* 21, 3 (March 2019), e12450.
[102] Weiser and Brown. 1995. Designing calm technology. *PowerGrid Journal* (1995).
[103] Wikimedia Foundation. 2019. API:REST API/Reference - MediaWiki. https://www.mediawiki.org/wiki/API:REST_API/Reference. Accessed: 2022-9-12.
[104] Wikipedia contributors. 2003. Wikipedia:Edit warring. https://en.wikipedia.org/w/index.php?title=Wikipedia:Edit_warring&oldid=1121114701. Accessed: NA-NA-NA.
[105] Wikipedia contributors. 2003. Wikipedia:Protection policy. https://en.wikipedia.org/wiki/Wikipedia:Protection_policy. Accessed: 2022-NA-NA.
[106] Wikipedia contributors. 2004. Category:Internet culture. https://en.wikipedia.org/w/index.php?title=Category:Internet_culture&oldid=1096814859. Accessed: 2022-NA-NA.
[107] Wikipedia contributors. 2004. Wikipedia:Dispute resolution. https://en.wikipedia.org/w/index.php?title=Wikipedia:Dispute_resolution&oldid=1125599763. Accessed: NA-NA-NA.
[108] Wikipedia contributors. 2005. Wikipedia:Biographies of living persons. https://en.wikipedia.org/w/index.php?title=Wikipedia:Biographies_of_living_persons&oldid=1125912067. Accessed: NA-NA-NA.
[109] Wikipedia Contributors. 2006. List of Wikipedias by edits per article. https://meta.wikimedia.org/wiki/List_of_Wikipedias_by_edits_per_article. Accessed: 2022-12-12.
[110] Wikipedia contributors. 2006. Wikipedia:Disruptive editing. https://en.wikipedia.org/w/index.php?title=Wikipedia:Disruptive_editing&oldid=1125728690. Accessed: NA-NA-NA.
[111] Wikipedia contributors. 2007. User:ClueBot NG. https://en.wikipedia.org/w/index.php?title=User:ClueBot_NG&oldid=391868393. Accessed: NA-NA-NA.
[112] Wikipedia contributors. 2009. Help:Range blocks. https://simple.wikipedia.org/w/index.php?title=Help:Range_blocks&oldid=6724630. Accessed: NA-NA-NA.
[113] Wikipedia contributors. 2010. Wikipedia:Wikipedia is anonymous. https://en.wikipedia.org/w/index.php?title=Wikipedia:Wikipedia_is_anonymous&oldid=1088618418. Accessed: NA-NA-NA.
[114] Wikipedia contributors. 2012. Wikipedia:Requests for page protection/Archive. https://en.wikipedia.org/w/index.php?title=Wikipedia:Requests_for_page_protection/Archive&oldid=1113712164. Accessed: NA-NA-NA.
[115] Wikipedia contributors. 2012. Wikipedia:Requests for page protection/Increase/Form. https://en.wikipedia.org/w/index.php?title=Wikipedia:Requests_for_page_protection/Increase/Form&oldid=1044821024. Accessed: NA-NA-NA.
[116] Wikipedia contributors. 2022. Wiki. https://en.wikipedia.org/w/index.php?title=Wiki&oldid=1109121394. Accessed: NA-NA-NA.
[117] Alex Yarovoy, Yiftach Nagar, Einat Minkov, and Ofer Arazy. 2020. Assessing the Contribution of Subject-matter Experts to Wikipedia. *Trans. Soc. Comput.* 3, 4 (Oct. 2020), 1–36.
[118] Eva Zangerle, Georg Schmidhammer, and Günther Specht. 2015. #Wikipedia on Twitter: analyzing tweets about Wikipedia. In *Proceedings of the 11th International Symposium on Open Collaboration* (San Francisco, California) *(OpenSym '15, Article 14)*. Association for Computing Machinery, New York, NY, USA, 1–8.
[119] Zhongheng Zhang, Hwa Jung Kim, Guillaume Lonjon, Yibing Zhu, et al. 2019. Balance diagnostics after propensity score matching. *Annals of translational medicine* 7, 1 (2019).
[120] Zheng, Mai, Gordon, and Nickerson. 2019. Bursty coordination in online communities. *ICIS 2019 Proceedings* (2019).
[121] Haiyi Zhu, Amy Zhang, Jiping He, Robert E Kraut, and Aniket Kittur. 2013. Effects of peer feedback on contribution: a field experiment in Wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2253–2262.


## A ROBUSTNESS CHECK

Received January 2023; revised April 2023; accepted May 2023

| | Ongoing Conflicts |
|---|---:|
| Total Articles | 3,807 |
| Protected Articles | 620 |
| Total Edits | 622,666 |
| Total Editors | 98,419 |

Table 10. Descriptive statistics for the set of articles in the Ongoing Conflicts category. This category of articles contains articles describing emerging hostilities between or within countries. We chose this category for a robustness check because it's a heavily edited emerging topic with a high percentage (16%) of protected articles.

| | Dropoff | | | Uptake | | | Retention | | | Difference | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Conditions | Mean(SD) | Min | Max | Mean(SD) | Min | Max | Mean(SD) | Min | Max | Mean(SD) | Min | Max |
| seven | 6.12(13.95) | 0 | 178 | 7.84(60.05) | 0 | 1077 | 2.06(4.72) | 1 | 65 | -1.72(56.51) | -1012 | 64 |
| fourteen | 7.5(15.98) | 0 | 173 | 10.63(79.38) | 0 | 1554 | 2.33(5.18) | 1 | 70 | -3.12(75.72) | -1493 | 137 |
| thirty | 9.96(24.03) | 0 | 391 | 14.98(102.2) | 0 | 2083 | 2.73(5.9) | 1 | 78 | -5.01(98.49) | -2023 | 324 |

Table 11. In line with our main findings for RQ2, we see that dropoff, uptake and difference have extremely large standard deviations. This suggests that page protection has inconsistent effects on the editor landscape. Specifically, we see standard deviations of nearly 100 editors for uptake and difference.

| | Conditions | **Gini** | | **Hoover** | | **20:20** | |
|---|---|---|---|---|---|---|---|
| | | Mean(SD) | p-val | Mean(SD) | p-val | Mean(SD) | p-val |
| **protected articles** | pre | 0.46(0.18) | – | 0.39(0.14) | – | 5.57(5.56) | – |
| | seven | 0.93(0.1) | 5.18E-280 | 0.9(0.14) | 0.00E-01 | 5.96(5.28) | 6.61E-43 |
| | fourteen | 0.92(0.1) | 4.28E-274 | 0.89(0.14) | 1.18E-299 | 6.31(5.26) | 4.68E-39 |
| | thirty | 0.91(0.1) | 1.96E-262 | 0.88(0.14) | 1.32E-282 | 6.87(6.35) | 7.82E-31 |

Table 12. Mirroring our main findings for RQ2, we see that all three measures significantly increase with time. Namely, we see the Gini and Hoover indices spike to nearly .9 immediately after page protection. This suggests that after page protection, nearly 90% of edits would need to be redistributed amongst contributors to achieve equality.

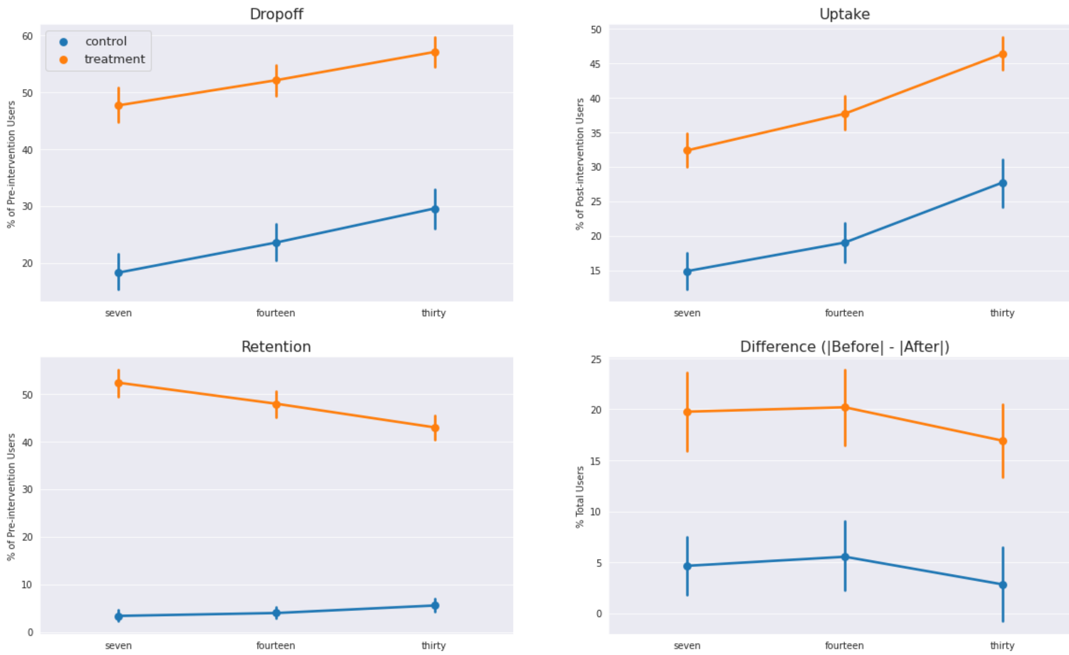


Fig. 6. Normalized participant engagement trends (Dropoff, Uptake, Retention, Difference) for protected articles (orange) and comparable unprotected articles (blue) in the Ongoing Conflicts Category. This robustness check furthers our findings that protected articles experience high user churn and inconsistent effects.

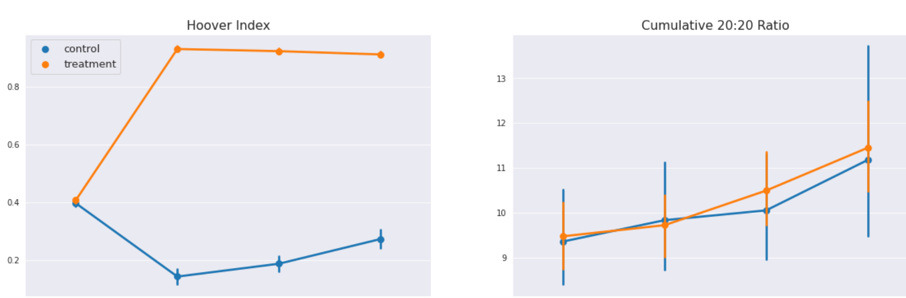


Fig. 7. Contributor concentration trends (Hoover Index and Cumulative 20:20 Ratio) for protected articles (orange) and comparable unprotected articles (blue) in the Ongoing Conflicts category. Consistent with our main findings, we see dramatic increases in the Hoover index. Error bars indicate 95% CIs.