Online Platform Policy and User Engagement

by

Qinglai He

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved May 2021 by the
Graduate Supervisory Committee:

Raghu Santanam, Co-Chair
Yili (Kevin) Hong, Co-Chair
Gordon Burtch

ARIZONA STATE UNIVERSITY

August 2021

ABSTRACT

Various activities move online in the era of the digital economy. Platform design and policy can heavily affect online user activities and result in many expected and unexpected consequences. In this dissertation, I conduct empirical studies on three types of online platforms to investigate the influence of their platform policy on their user engagement and associated outcomes. Specifically, in Study 1, I focus on goal-directed platforms and study how the introduction of the mobile channel affects users' goal pursuit engagement and persistence. In Study 2, I focus on social media and online communities. I study the introduction of machine-powered platform regulation and its impacts on volunteer moderators' engagement. In Study 3, I focus on online political discourse forums and examine the role of identity declaration in user participation and polarization in the subsequent political discourse. Overall, my results highlight how various platform policies shape user behavior. Implications on multi-channel adoption, human-machine collaborative platform governance, and online political polarization research are discussed.

ACKNOWLEDGMENTS

Pursuing a PhD had become my dream since my junior year in college. The last five years were not easy but still full of joy and growth. I greatly appreciate my life opens the door for me and brings me here.

First, I want to thank my parents. Your love means the whole world to me whenever I face any challenges in my life. Thank you for selflessly supporting my dream and my decisions. I love you, and I will also give you the best in my life.

I also want to thank my incredible committee, Raghu Santanam, Kevin Hong, and Gordon Burtch, who guided me throughout my PhD study. You earn all credits for all my achievements today. I was extremely fortunate to be your student and get the best mentorship. Thank you for all conversations during and after work, being tolerant of my mistakes, giving me opportunities and unlimited supports in the critical time. I cannot be who I am today without your mentorship.

I want to thank fantastic professors who taught or influenced me in the last five years: Victor Benjamin, Pei-yu Chen, Sang Pil Han, Ben Shao, John Zhang, Dokyun Lee, Sungho Park, Tim Richards, Glenn Hoetker, Fernando Leiva Bertran, Seung Ahn, Andrea Morales, Kathleen Moser. You taught me valuable research skills and showed me how a great researcher and teacher are supposed to be. I appreciate your passion, attitude, and thoughts you devoted to the work.

Also, I want to thank Angelina Saric and my PhD cohort, Amin, Chen, Jingbo, Kumar, Xueyan, Xiaohui, Ying, Ziru. Thanks for your company and care. You taught me a lot and helped me grow. My PhD life was much better because of you.

I want to thank my best friend, Xiao Ting. You have supported me for more than twelve years. You always believe in me and encourage me to be a better version of myself when I experience tough times. Thank you so much for all your patience and understanding.

I also want to thank all my friends and everyone I met in the past five years, especially Shuhan, Brenden, Shambam, Vandith, Chris, Lingyan, Li, Qin, Xiaoxiao, Andrez, Michael, Sai, Nishanth, and Zhouxuan. I will never forget the time we had together. I also sincerely wish you a happy life ahead.

I want to thank Erasmus University Rotterdam, the University of Wisconsin – Madison, and the University of Rochester for inviting me for a job talk. These opportunities mean a lot to me. It was fantastic to meet all your amazing faculties.

Finally, I want to share some of my current thoughts with anyone who is reading this acknowledgment. My PhD study enables me to acquire the knowledge I desired and helps me know the world better. Meanwhile, I am confused and sometimes lost when facing different voices, values, and conflicting issues. At the personal level, I still need to learn how to handle friendship, love, career, life, society, and the world better. I want to tell myself to stay honest and sincere to the world. Be yourself but also adapt to the environment. Enjoy the process, work hard and be creative. When I first faced my career choice, I told my mom that I wanted to do something beneficial for the world. I still have that expectation. I cannot guarantee, but I wish I would not give up trying.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

With the rapid development of technology and increasing trend of extending real-world activities into the virtual, online platforms have become the main venue for various activities such as learning (Huang et al. 2021; Santhanam et al. 2016), networking (Li et al. 2017; Garg et al. 2018), and political discussion (Bail et al. 2018; Levy 2021). User engagement drives platform growth and, ultimately, the revenue increase. Therefore, online platforms apply various strategies and platform policies to stimulate user engagement and adapt to the fast-changing technological and business environment.

With the focus on user engagement, in this dissertation, I conduct a series of empirical studies on three types of emerging platforms. The first online platform I study is goal-directed platforms. Goal-directed platforms have experienced rising popularity to assist individuals' goal pursuit in various aspects of life, such as financial management, weight loss, and skill learning. Prior literature has demonstrated the importance of diverse approaches to individuals' goal pursuit activities. However, few studies have investigated how technology-mediated goal pursuits affect individuals' behaviors. In Study 1, I perform a series of empirical analyses to examine the impacts of multi-channel adoption on goal pursuit activity and persistence. The results indicate that mobile adoption improves overall goal pursuit effort by 140.1%. A positive effect on goal pursuit persistence is also observed. Most notably, the enlarging gap between different types of students has been found when a new technological channel is introduced. Particularly,

users with high-level goal specificity and high learning competency achieve more considerable improvement from adopting the mobile channel.

In addition to motivating user participation and performance, as platforms play a more critical role in our daily activities, the need for regulating online content has grown exponentially. Volunteer moderators are given the role to help maintain a healthy online environment, and they have become the growing and special user group on platforms, particularly social media. As a result, volunteer moderators have been the essential workforce for platform governance.

However, human moderation suffers from a limited capacity in moderating massive and undesirable content. As platforms move toward the technical and automated mode of governance, there is a growing concern over de-humanization and whether machines would lead volunteer moderators to reduce their contributions. To understand the role of these increasingly popular bot moderators, in Study 2, I conduct an empirical study to examine the impact of machine-powered regulations on volunteer moderators' behaviors. With data collected from 156 subreddits on Reddit, a large global online community, I found that delegating moderation to machines augments volunteer moderators' role as community managers. Human moderators engage in more moderation-related activities, including 20.2% more corrective and 14.9% supportive activities with their community members. Importantly, the effect manifests primarily among communities with large user bases and detailed guidelines, suggesting that community needs for moderation are the key factors driving more voluntary contributions in the presence of bot moderators.

Lastly, I turn my attention to online political discussion forums and investigate the influences of identity declaration on user participation and polarization in subsequent political discourses. Political identity has become a critical social identity in the era of digital platforms. Literature has examined identity disclosure in numerous online platforms. However, little attention has been paid to political identity and its impact on online political discussion. Our study takes advantage of a policy change on Reddit and utilizes exogenous shock to study how political stance disclosure causally impacts subsequent political discourse. Our results suggest an important trade-off between user interaction and polarization. Specifically, identity declaration stimulates the idea exchange between different political perspectives. However, such interactions also become more polarized and partisanship. These results highlight an important trade-off of identity declaration in managing online political discourses. We further reveal the underlying mechanism from aspects, including discourse type and participant political stance. Managerial implications are also discussed.

The rest of the dissertation is as follows. Chapter 2 describes the research detail of Study 1, followed by Study 2 and Study 3 in Chapter 3 and Chapter 4, respectively. Finally, I summarize the main findings of each study and conclude the whole dissertation in Chapter 5.

CHAPTER 2

CLOSING THE GAP OR WIDENING THE CHASM: IMPACT OF MOBILE CHANNEL ADOPTION IN GOAL-DIRECTED PLATFORMS

2.1 Introduction

With the rising trend in transforming activities from the physical world to the virtual, individuals and organizations are increasingly utilizing their capabilities for self-improvement. Many recent products and services are designed with the intent to help individuals set and track various types of goals. There are several goal-directed software applications, web platforms, wearable devices, mobile ecosystems, IoT devices, and more for helping individuals meet personal goals spanning health, personal finance, education, lifestyle, and others. The market appetite for goal-directed platforms has grown significantly over the years as evidenced by high-profile market transactions, such as Under Armour's $475 million-dollar acquisition of MyFitnessPal, a goal-directed platform that focuses on exercise and nutrition.[1]

Despite the market growth for goal-directed platforms in recent years, few works have examined the impact of the introduction of technology on goal pursuit. While some analogs to these questions have been explored in information systems research (Xu et al. 2017; Liu et al. 2016; Jung et al. 2019), goal-directed platforms are different from other types of online platforms. Users' engagement on the goal pursuit platforms is a series of efforts towards achieving the goal (Fishbach and Ferguson 2007; Smith et al. 1990). This type of purposeful interaction with meaningful progression towards an end goal is very

[1] https://www.wsj.com/articles/under-armour-to-acquire-myfitnesspal-for-475-million-1423086478

different from other types of online platforms where usage is intended for discrete purposes rather than a cumulative effort for progressing towards some end goal. Therefore, compared to other platforms, a *regular* and *continuous* engagement on goal-directed platform has become an essential factor to users' success (Swann and Rosenbaum 2018) and the main interest of platform operators.

Most importantly, technology's impact on the ability of different user groups to pursue goals may also vary based on some goal pursuit characteristics. Prior literature has highlighted the essential role of goal-related factors, such as goal specificity and user competency, in individuals' goal achievement (Fishbach and Ferguson 2007; Liu et al. 2016; Locke 1996; Redding 2014; Schunk 1991). However, extant multi-channel adoption literature focusing on other contexts provides very little empirical evidence on these angles. From the managerial perspective, it is also critical for platform operators to understand how different users benefit from adopting an additional technological channel and then apply appropriate strategies to manage channel introduction and help users succeed.

This study attempts to address the above described research gaps through an empirical study investigating the impact of multi-channel adoption on users' goal pursuit activities within a goal-directed platform. Specifically, we focus on goal pursuit effort and persistence as two important dimensions for goal achievement (Huang et al. 2016; Liu et al. 2016). To disentangle the impacts of technology adoption, we also consider users' heterogeneity in two goal pursuit characteristics (i.e., goal specificity and goal pursuit competency). Formally, we seek to address the following research questions in this research:

- Does multi-channel adoption improve goal pursuit effort and persistence?
- What goal pursuit characteristics determine the ability of users to benefit differently from adopting the multi-channel?

To operationalize this research, we collaborated with Picmonic, a leading goal-directed platform in the online education space based in the United States. Picmonic provides visual learning tools and systems that help students effectively learn medical courses and prepare for standardized exams. The platform offers multiple activities such as video-based learning and self-assessment activities across various domain topics so that users are able to pursue any combination of these activities to reach their goals. The acquired dataset covers all user activities during August 2017 and August 2018. During this time frame, a portion of users initially started with traditional PC-based experience, but later they adopted mobile channel. Following the best practices in the literature (Xu et al. 2017; Jung et al. 2019). We constructed a dataset containing user activities both before and after their mobile adoption, and then use propensity score matching (PSM) to identify and match comparable mobile adopters and non-adopters. We then estimate the effect of multi-channel adoption with a difference-in-differences (DID) model.

Results from the PSM-DID model suggest that multi-channel adoption improves overall goal pursuit effort by 140.1%. A positive effect on goal pursuit persistence is also observed. Specifically, users spent 0.656 days learning content after they adopted mobile channel. The increases are observed in both knowledge learning and knowledge testing activities, but users shift their attention to knowledge testing as it aligns best with the affordances offered by the mobile channel. Interestingly, user heterogeneity leads to differential user benefit realization from adopting multi-channel. Users with high-level

goal specificity are observed to spend 0.234 more days and 56.5% more effort in their goal pursuit than users with a less specific goal. Robustness checks and replication of data analyses further confirm that the results are robust under various scenarios.

Our research makes three key contributions. First, we contribute to goal pursuit literature (Fishbach and Finkelstein 2012; Uetake and Yang 2018) by focusing on the fast-growing but underemphasized area of technology-mediated goal pursuit. Particularly, we study the positive effects of mobile channel adoption on goal pursuit effort and persistence. Second, our research also contributes to multi-channel adoption literature (Xu et al. 2017; Liu et al. 2016) by introducing the theoretical perspective of goal pursuit and adding empirical evidence in the context of goal-directed platforms. Our results suggest that the heterogeneous goal specificity and goal pursuit competency moderate the impacts of multi-channel adoption on users' following activities. Third, from a practitioner perspective, our findings illustrate how channel introduction can serve as a strategy to stimulate users' goal pursuit. Our study highlights that adopting additional channel leads to the shift in users' goal pursuit activities. It is necessary for platforms to simultaneously maintain multiple channels and take appropriate interventions to assist users who benefit less from mobile application adoption.

2.2 Literature Review and Hypothesis Development

To aid in the hypotheses development for this study, we first review several key constructs and relevant mechanisms from goal pursuit theory. With an understanding of the foundational concepts, we discuss how mobile adoption would impact goal pursuit, considering goal pursuit effort and persistence. Next, we differentiate users based on their

goal specificity and goal pursuit competency. We then discuss how users with different attributes would benefit disproportionally from adopting the multi-channel channel.

2.2.1 Goal Pursuit Theory

Goals are a driving factor in human decisions. According to the definition by Fishbach and Dhar (2005), goals are cognitive structures that can be represented in terms of movement and progress towards some *abstract* or *specific* end state. The study of goal pursuit has been of interest to researchers for decades. Over time, goal pursuit theory has emerged as a vehicle for exploring individuals' ability to achieve goals (Locke 1996; Fishbach and Finkelstein 2012; Jiao and Cole 2015).

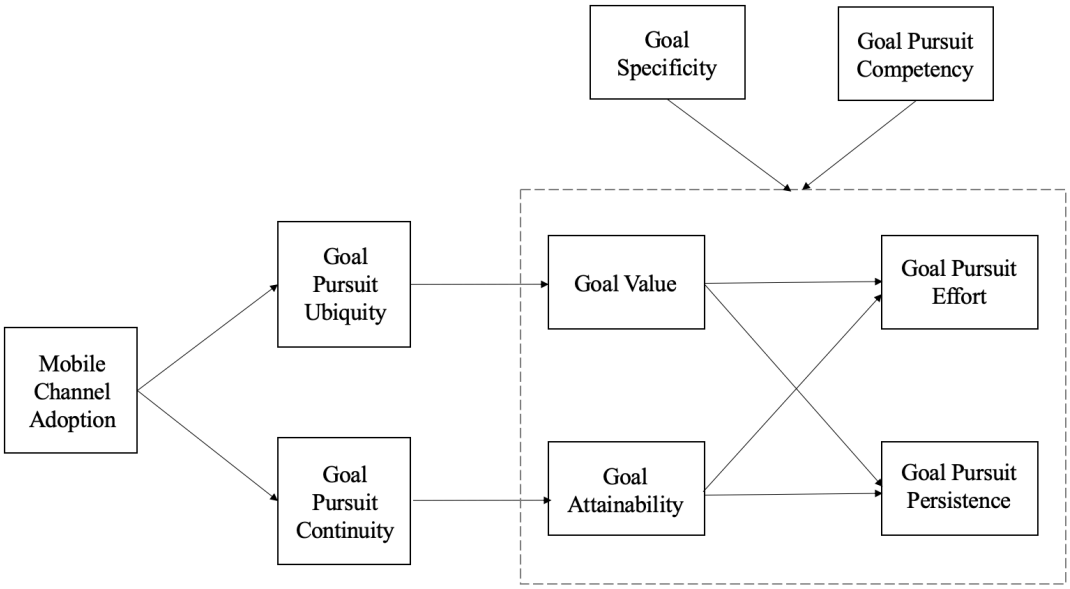


Figure 1. Key Constructs and Research Framework

In this study, we focus on two essential constructs from goal pursuit theory: (a) goal pursuit effort, and (b) goal pursuit persistence. Goal pursuit effort refers to the amount of effort individuals expend to accomplish their goal (Locke 1996; Feather 1962),

whereas goal pursuit persistence indicates how often or how long an individual maintains their effort over time. These two constructs reveal *the magnitude and persistence of effort* and taken together complement our understanding of goal pursuit outcomes from both effort and time dimensions.

Extant research on individuals' motivation suggests that goal pursuit effort and persistence are driven by the goal's *value* and *attainability* (Fishbach and Finkelstein 2012; Brandstätter and Frank 2002; Huang and Zhang 2011). When individuals perceive a goal to be more desirable (i.e., a high value of a goal) and achievable (i.e., high attainability), they are likely to exert more effort to attain the goal and continue this effort for a longer time. Therefore, when it comes to the effect of a particular factor on eventual goal pursuit activities, it would be essential to consider its impact on these determinant constructs (i.e., goal's value and attainability).

A number of factors could affect an individual's perception of their goal's value and attainability. When imagining a goal, individuals will do so at varying levels of specificity. *Goal specificity* is the degree of precision with which a goal is specified (Locke et al. 1981; Locke 1996; Fishbach and Ferguson 2007). Individuals are able to describe a goal by specifying the scope of the goal they want to attain, how they will achieve it, and in what time span they will accomplish it. For example, achieving a good score on an exam can be described as a nonspecific goal (e.g., "Do your best"), whereas a more specific goal would include an explicit score threshold to pass (e.g., "Get at least 90 percent"). Overall, goal specificity will impact individuals' ability to pursue their goal, and ultimately, their ability to achieve the goal (Smith et al. 1990; Klein et al. 1990; Wallace and Etkin 2018; Künsting et al. 2011).

In addition to goal specificity, *goal pursuit competency* is another important factor in goal achievement (Redding 2014; Schunk 1991). In general, two facets account for differences in *competency* levels among individuals: (a) knowledge about the goal; and (b) strategic usage of the resources for goal pursuit. Given goal pursuit is a continuous process, the competency formed in the earlier stages of goal pursuit would impact individuals' future goal achievement. Therefore, it is important to consider the difference in individuals' goal pursuit competency when we study goal achievement.

Goal pursuit theory has been utilized extensively in traditional research contexts. In recent years, the ubiquity of information technologies has given rise to goal-directed platforms that aid with financial management, fitness, and skill-learning; this development has attracted researchers' attention (Uetake and Yang 2018). Overall, use of technology has become a core part of goal tracking and pursuit for many individuals. However, despite this trend, prior literature mainly focuses on users' goal pursuit behaviors irrespective of associated technological channel. Little research has investigated the introduction of technology during the goal pursuit process and examined its effects on overall goal pursuit effort and persistence.

To address this research gap, we focus on the intersection of goal pursuit and mobile channel adoption through the lens of goal pursuit theory. Focus is placed on mobile channel adoption due to its common usage for goal completion. Further, given the impact of goal pursuit characteristics, we examine the moderating effect of various goal pursuit characteristics including goal specificity and goal pursuit competency. We propose our hypotheses in the following four subsections.

2.2.2 Mobile Channel Adoption and Goal Pursuit

Technology adoption is a common research stream across many domains. In particular, adoption of mobile devices is a frequent focus of scholarly pursuit. Within business disciplines, there is ample evidence demonstrating that mobile channel adoption has a positive impact on a variety of business outcomes, such as sales (Xu et al. 2017; Huang et al. 2016), service demand (Liu et al. 2016), and user engagement (Lee et al. 2017; Jung et al. 2019; Son et al. 2016). Overall, prior research suggests the capabilities provided by the mobile channel and how they influence users' behaviors and expectations.

There are two capabilities provided by a mobile channel that are lacking within a traditional PC channel (i.e., PC-based web browser): *ubiquity* and *continuity* (Jung et al. 2019; Xu et al. 2017). First, increased Internet access afforded by the mobile channel enables users to access online platforms outside of the location and time constraints that bound the traditional PC channel (Chae and Kim 2004; Orr 2010; Bang et al. 2013 a; Jung et al. 2019). Users can reach goal-directed platforms more consistently as a result of the increased ubiquity. Mobile channel adoption further helps users on a variety of goal pursuit relevant activities such as planning new goal activities, continuing prior activities, and tracking overall goal progress. Instances where users are not able to pursue goal activities due to space and time constraints imposed by the traditional channel are no longer present when using the mobile channel.

Second, along with enhanced ubiquity, the mobile channel can further improve the *continuity* of a users' goal pursuit. In contrast to the experience wherein users perform all goal-pursuit activities on a single channel (i.e., PC-based channel), with mobile devices, users are able to continue their previous activities (e.g., unfinished sessions on the PC

channel) by interchangeably using different channels as needed. Therefore, temporary changes of location or time constraint are less likely to suspend the flow of users' goal pursuit activities. Users who adopted mobile channel are able to more frequently access the content when compared to those with access to only a PC-based channel.

Enhanced accessibility and continuity provided by mobile channel affect the *value* and *attainability* of goals. In the context of this study, greater accessibility introduced by a new channel suggests an additional approach towards the goal. The presence of multiple (rather than fewer) channels to attain a goal augments the perceived value of the goal (Higgins 2000; Higgins et al. 2003; Kruglanski et al. 2011). Further, goal achievement becomes more feasible after mobile adoption. Compared to users who only use a PC-based application, those who adopt the mobile channel will possess a greater capacity and autonomy to access the goal-directed platform. Mobile enables users to better plan and manage goal activities, leading to increases in effort investment (more goal pursuit effort) and time spent (more goal pursuit persistence) (Zhang 2008). Therefore, we propose the following hypothesis:

H1a: Adoption of the mobile channel enhances users' goal pursuit effort.

H1b: Adoption of the mobile channel increases users' goal pursuit persistence.

2.2.3 Heterogeneity in Users' Goal Specificity

A stream of literature on goal pursuit has provided evidence showing the positive impact of goal specificity (Wallace and Etkin 2018; Künsting et al. 2011). For example, Tubbs (1986) found that goals incorporating specific requirements or standards are likely to increase individuals' ability to self-evaluate and further boost their goal performance.

Within problem-solving scenarios, Künsting et al. (2011) also indicated that individuals with the assignment of a specific goal achieve better performance.

Three pathways may explain why users with a highly specific goal achieve better performance (Locke 1981): goal uncertainty, strategic usage of goal pursuit activities, and timely feedback. First, a high level of specificity will enhance the interpretation and evaluation of a goal. Individuals with high goal specificity will have less uncertainty about what goal they are expected to achieve and what actions must be taken to accomplish their goal. This increased certainty would lead individuals to pursue their goal with less concern of potential obstacles that could inhibit goal progress (Locke 1981; Klein et al. 1990; Tubbs 1986). Second, a specific goal will lead individuals to have a more careful consideration of explicit objectives and how to approach them strategically. When multiple approaches toward a goal are available, it would be easier for users with a specific goal to compare the effectiveness of different approaches and choose the best one to assist their goal pursuit. Third, goals with high specificity would help individuals better measure the distance between their current position and their end state (Campion and Lord 1982). Individuals can then adjust their activities to pursue their goals more effectively. These behaviors are unique to goal-seeking behavior and diverge from previous studies examining IT adoption outside of goal pursuit.

It is important to consider the context of each individual user's level of goal specificity, as adoption of the mobile channel may induce an even larger gap between those with high-specificity and low-specificity. First, ubiquity and continuity allow users to reduce the uncertainty of goal achievement as the mobile channel provides more opportunities for users to continuously perform goal pursuit activities. Second, the

availability of the mobile channel enables diversification of the actions that individuals take for goal pursuit; strategic goal pursuit thus becomes possible. Third, mobile channel also enables users to receive timely feedback with fewer space and time constraints. With a highly accessible mobile channel, it becomes easier for users to review their goal progress and adjust their activities accordingly (Fishbach and Finkelstein 2012). In sum, those with high levels of goal specificity can derive more benefits from the mobile channel. Bearing the above discussion in mind, we hypothesize that:

H2: Compared to users with low-level of goal specificity, the positive impact of mobile adoption on goal pursuit effort and persistence is stronger for users with high-level goal specificity.

2.2.4 Heterogeneity in Users' Goal Pursuit Competency

Competency is the ever-evolving accumulation of knowledge and skills that facilitate learning and other forms of goal attainment (Redding 2014). Individuals' competency is improved along with the growth of their knowledge and experience in goal pursuit. For example, college students enhance their competency as they progress on the academic path from freshmen year study to senior year study (Redding 2014).

According to the definition of competency in the goal pursuit context, the degree of users' goal pursuit competency is mainly driven by individuals' knowledge and experience (Redding 2014). We propose that accumulation of knowledge and experience would allow adopters of mobile channel to pursue their goals with more effort and persistence.

First, individuals with high goal competency have more comprehensive knowledge and understanding of the objective they are trying to accomplish. Similar to the impacts of goal specificity, the knowledge these users possess will reduce the uncertainty of goal pursuit and even help them identify the inconsistency between their desired objective and the actual outcome of their actions (Sitzmann and Yeo 2013). When the mobile channel is introduced, high competency users are more likely to better evaluate the merits of different channels and apply them more effectively.

Most importantly, individuals with high competency usually accumulate more goal pursuit experience (Redding 2014). Thus, they are more likely to learn how to manage their future goal pursuit better, including learning lessons, skills, and strategies. These skills and experience can enable users to know their environment better so that they can achieve their goals strategically (Schunk 1991). Therefore, after adopting the mobile channel, individuals with high competency are less likely to get distracted, demotivated by setbacks, or experience conflict from mobile channel adoption. Therefore, more goal pursuit effort and persistence are expected for high competency individuals (Locke 1996; Blumenfeld 1992; Schunk 2003).

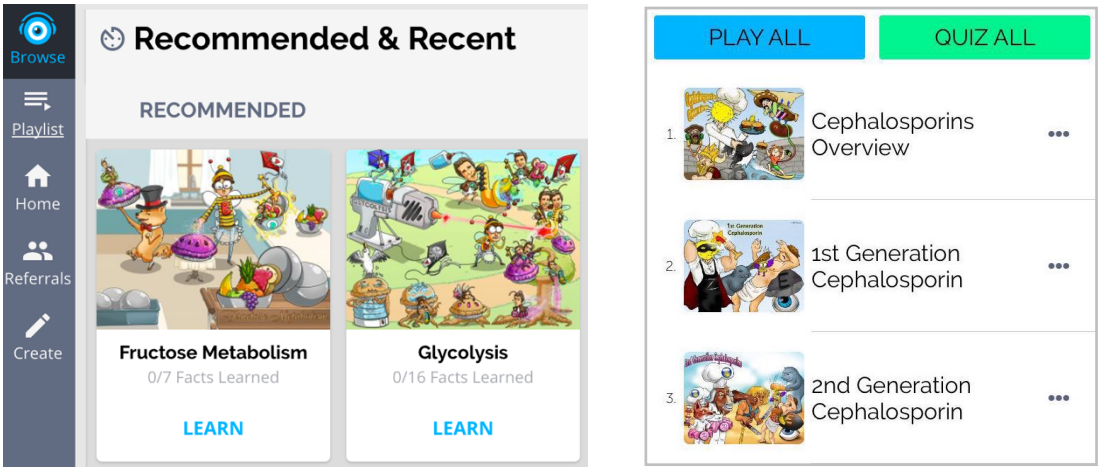
Given that high goal-competency users are more strategic in their goal pursuit, we expect that adopting the mobile channel will stimulate them to expend more effort and be more persistent. Therefore, we propose our last hypothesis:

H3: Compared to users with low goal-competency, the positive impact of mobile adoption on goal pursuit effort and persistence is stronger for users with high goal pursuit competency.

2.3 Research Context

2.3.1 Data Source

To examine how mobile adoption impacts users' goal pursuit, we utilize an online learning platform as the empirical context of this study. Online learning platforms have emerged as one of the most popular online platforms. They are also representative of what a typical goal-directed platform seeks to accomplish; that is, the primary objective of these platforms is to help their users achieve an education-oriented goal.



(a) Website (b) Mobile App

Figure 2. Screenshots of the Webpages of Picmonic

We conduct our study in collaboration with Picmonic, a sizeable online learning platform in the United States. Picmonic's primary user group consists of students pursuing careers in healthcare (e.g., medical, nursing, etc.). They provide video-based courses that allow users to learn content visually. Figure 2 showcases the interface of Picmonic. On the left, learning materials are organized into various short-session mnemonic content "cards." Cards are considered as the basic learning unit within the

Picmonic platform. Several cards can be strung together in a "playlist." For example, users can choose a playlist called *anatomy* to queue several anatomy-related cards for viewing. Playlists are further organized into different pathways such as *body system* and *courses*, and users can choose to follow the pathways to guide their learning. Picmonic also offers quizzes corresponding to the knowledge contained within each card. Students gain knowledge relevant to their field by learning from the mnemonic devices on each card and taking quizzes.

Picmonic launched with a traditional web presence in 2013 and later released a mobile channel in February 2015. Picmonic's mobile channel retains its core features but faces similar constraints that all mobile channel share, such as limited screen size. Users are able to access all content provided by the platform through the mobile channel to perform goal pursuit activities. In this study, we focus on learning from Picmonic cards and quizzes as these two goal pursuit activities are available to users on both the traditional PC-based and the mobile channel.

2.3.2 Sample and Variables

The Picmonic dataset spans from August 2017 to August 2018. At the beginning of the study period, Picmonic's mobile channel had been released for two years (i.e., it was released in February 2015). One advantage offered by this dataset is that since the mobile channel had already been released for some time, it had matured with stable functionality and features. Further, any potential bias generated by early adopters is eliminated by using a data sample that starts two years after the release of the mobile app. Thus, the

confounding impact of users' expectations on mobile application availability is likely not a concern (Caliendo and Kopeinig 2008).

From August 2017 to August 2018, 116,841 new users registered an account on Picmonic including free-trial users and subscribers. During this period, 83.1% of new user account registration occurred through the PC-based application. Thus, PC is still the dominant channel for user acquisition. Among those who registered through the PC channel, 13.4% of them adopted the mobile channel afterward.[2] Of these users, 30.3% of users adopted the mobile channel within a week after creating their account. Moreover, 48.4% of users adopted mobile within three weeks of their initial registration date.

Four categories of data are utilized in this research: (1) user registration data; (2) users' card learning behaviors; (3) quiz records; (4) premium service subscription records.[3] Before analyzing the data, a series of data preprocessing steps were completed to create a clean and structured dataset for analysis, described as follows:

- First, the data provided by Picmonic contains account information of students, instructors, administrators, and others. We focus on student accounts in this study as they represent more than 97% of users on the platform. Students are also the target users of the platform.
- Further, given our identification strategy (described in Section 4), the user sample utilized in this study consists only of users who *registered* their account using the traditional PC channel as these users have a more clearly defined pre-adoption

[2] For users who created their account on the mobile channel, nearly 10.1% adopted the PC channel afterward. Among all users in our observational window, 8.4% of them adopted both PC and the mobile channel in total.
[3] Picmonic produces learning materials and relevant quizzes for users, but it does not provide the actual standardized test. Therefore, the final exam score is not available to us.

period (only on the PC channel) and post-adoption period (on both PC and the mobile channel).

- Lastly, Picmonic's free-trial users can only access a small amount of content every day with the system providing a natural limit to goal-pursuit activities. Thus, we only include users who subscribed to the premium service of Picmonic into our sample to ensure users are more homogeneous in their motivation and ability to pursue goal-related activities.

After completing data pre-processing steps, a subset of 7,935 users is identified for analysis, including 1,839 mobile adopters and 5,996 non-adopters. From this subset, a series of user-level variables are generated, such as users' tenure, playlist creation, learning activities, and subscription status. Detailed definitions of variables and summary statistics are shown in Table 1. Note that in Table 1, subscript *i* denotes each user. All summary statistics presented in Table 1 are representative of the raw data values provided in the Picmonic dataset.

2.4 Research Methodology

2.4.1 Empirical Strategy

A quasi-natural experiment approach is performed in this study to examine the causality of mobile adoption on users' goal pursuit. It would be ideal to investigate our research question through a randomized experiment by randomly assigning users to adopt the mobile channel (i.e., treatment group) while other users are not (i.e., control group). The random assignment procedure would enable a causal evaluation of mobile adoption's impact on goal pursuit by simply comparing the performance between the treatment

group and the control group. However, such an approach is practically impossible to implement in this context since the mobile channel is already publicly available to all users. Therefore, following prior research that examines mobile channel adoption, we take a quasi-natural experiment approach (Jung et al. 2019).

Table 1. Definition and Summary Statistic of Mobile Channel Adopters

Variable & Measurement	Definition	Mean	S.D.	Min	Max	N
$Tenure\_Wk_i$	The number of weeks elapsed after utser *i* registered on the platform	10.899	17.705	0	98	1,839
$Has\_Playlist_i$	Dummy variable. 1 if user *i* has created a playlist on the platform; 0, otherwise.	0.574	0.494	0	1	1,839
Cur_Paid_i	Dummy variable. 1 if user *i* subscribed to mobile app at the time of adopting mobile app; 0, otherwise.	0.898	0.303	0	1	1,839
$Num\_Card\_Pre_i$	The cumulative number of cards user *i* had attempted till adopting mobile app.	84.835	292.792	0	5,661	1,839
$Num\_Quiz\_Pre_i$	The cumulative number of quizzes user *i* had taken till adopting mobile app.	37.850	135.742	0	2,889	1,839
$Card\_Day\_Pre_i$	The cumulative number of days user *i* had attempted cards till adopting mobile.	6.908	16.297	0	203	1,839
$Quiz\_Day\_Pre_i$	The cumulative number of days user *i* had taken quizzes till adopting mobile.	5.221	12.521	0	175	1,839
$Num\_Card\_2w_i$	The number of cards user *i* had attempted during two weeks leading to the adoption of the mobile app.	20.999	66.950	0	1,559	1,839

The adoption of Picmonic's mobile channel is used to operationalize the quasi-natural experiment approach taken in this research. There are a large number of users

who first sign-up with Picmonic through the traditional PC channel, and only later adopt the mobile app after some time of using the platform. We consider user adoption of the mobile channel as a treatment, and subsequently, mobile adopters are the treatment group of interest in this study. The date on which a user first adopts the mobile channel is identified as treatment assignment time (Xu et al. 2017; Jung et al. 2019). We identify a control group from the non-adopters by applying propensity score matching (PSM) to match mobile adopters and non-adopters who share similar characteristics. Further, we estimate the effect of mobile adoption on users' goal pursuit by comparing the behaviors of the treated users versus the control users in the post-treatment period, relative to their behavior differences in the pre-treatment period. Precisely, the effect of the treatment is estimated using a difference-in-differences (DID) model. Both of the PSM and DID model analyses will be detailed in the following sections.

2.4.2 Econometric Analysis

2.4.2.1 Propensity Score Matching

One of the most critical processes of the quasi-natural experiment approach is to construct a high-quality matching between treatment and control groups. Here we use PSM as our primary matching method for pairing mobile adopters and non-adopters based on the observable variables in the pre-adoption period (Rubin 2006; Caliendo and Kopeinig 2008). The matching process is performed as follows:

- First, to ensure each sample has at least two periods of pre-adoption observation for the estimation in the DID models, given our panel data is organized in bi-weekly intervals, we focus on adopters who registered on the platform for at least three

weeks before adopting the mobile channel. The same approach has been adopted by prior literature (Jung et al. 2019).

- Second, for each adopter, we identified a set of non-adopters that registered with Picmonic during the same week as the adopter. Next, a series of measures based on non-adopter and adopter usage behaviors are generated for matching. The matching process considers only data points present during the time span of pre-adoption for the eventual mobile adopter (Caliendo and Kopeinig 2008). The variables used in matching include users' behavioral features (e.g., $Tenure\_Wk_i$ and $Has\_Playlist_i$) and also varieties of cumulative learning activities (e.g., $Num\_Card\_Pre_i$, $Num\_Quiz\_Pre_i$, $Card\_Day\_Pre_i$, and $Quiz\_Day\_Pre_i$) prior to actual (for treated users who adopted the mobile channel) or hypothetical (for the highly similar control users who did not adopt the mobile channel) mobile adoption date. These measures of cumulative learning activities provide information regarding users' goal pursuit progress, diversity in goal pursuit means, and frequency of actions towards goal pursuit. Prior studies (Fishbach and Finklstein 2012; Huang et al. 2011) also suggest that these factors play an important role in subsequent individuals' strategic behavior (e.g., mobile channel adoption in this study) as well as goal pursuit.
- Third, given that card learning is the most popular activity launched by Picmonic, we also extracted users' card learning effort (i.e., *Num_Cards_2w*) for the two weeks prior to their mobile adoption to control the extent of their recent activities. Compared to measures of cumulative learning activities, *Num_Cards_2w* indicates the momentum towards the adoption of mobile channel as well as users

subsequent goal pursuit. Given the content accessibility between premium service subscribers and non-subscribers is different, we also generated Cur_Paid_i to indicate users' subscription status at the time of adopting the mobile channel. We included this variable in the PSM to produce more meaningful matches.

- Finally, one-to-one without replacement PSM with a caliper of 0.05 is performed using the aforementioned measures. Since users adopted the mobile channel at different time points across the observation window, we applied a stratified matching approach to achieve high matching performance (Caliendo and Kopeinig 2008). Specifically, we performed the matching procedure for each week on which mobile adoption occurred and then aggregated all weekly matching results as the final matched sample in the following data analyses. In all, 762 matched pairs are generated and used in the following DID analyses.

Note that the mobile application was released before our observational period. In this study, the mobile app was always available; hence there is no concern that users would change behavior in anticipation of an impending release of a mobile app. In other words, users' actions in the pre-adoption period are less likely affected by mobile channel adoption.

To evaluate the PSM-based matching approach, we compare all measures used in the PSM between the treatment and control group after matching using a *t*-test. We present our results in Table 2. After conducting matching, there are no statistically significant differences between the treatment and control group in terms of all matching variables. This result demonstrates that the matched groups are quite similar in terms of their observed characteristics and pre-treatment platform behaviors.

Table 2. T-test of Variables of Treatment and Control Group After Matching

N	Variable	Mean (Control)	Mean (Treated)	*t*-value	*p*-value
1,524	Tenure_Wk$_i$	20.566 (0. 718)	19.797 (0. 665)	0.786	0.432
1,524	Has_Playlist$_i$	0. 559 (0. 018)	0. 589 (0. 018)	-1.191	0. 234
1,524	Num_Card_Pre$_i$	3.522 (0.065)	3.587 (0. 067)	-0.694	0.488
1,524	Num_Quiz_Pre$_i$	2.771 (0. 063)	2.791 (0.064)	-0.217	0.828
1,524	Card_Day_Pre$_i$	1.946 (0.038)	2.002 (0.039)	-1.029	0.304
1,524	Quiz_Day_Pre$_i$	1.719 (0.040)	1.743 (0.039)	-0.416	0.678
1,524	Num_Card_2w$_i$	1.603 (0.066)	1.611 (0.066)	-0.091	0.927
1,524	Cur_Paid$_i$	0. 902 (0. 011)	0.903 (0. 011)	-0.086	0.931

Notes: (1) *Num_Card*, *Num_Quiz*, *Num_Card_2w*, *Card_Day* and *Quiz_Day* are log-transformed; (2) Caliper of 0.05 is used to generate the matched pairs.

2.4.2.2 Difference-in-differences Method

To analyze the effects of multi-channel adoption, a DID model is performed using 7,620 biweekly observations extracted for 1,524 matched users. Each user has five periods of observation, consisting of two periods of pre-adoption, one period when the treatment occurs, and two additional periods of post-adoption observation. In the DID model, the matched non-adopters are scrutinized with the same pre- and post-adoption observational windows as their matched adopting user. Additionally, a new variable, $After_{it}$, is generated and added to the panel dataset to represent whether the observation happened before or after mobile adoption.

To test H1, a DID model is constructed (see Model (1)). In this model, the dependent variable $Activity_{it}$ is defined as the goal pursuit effort and persistence for all users *i* in month *t*. Goal pursuit effort is measured by Num_Card_i and Num_Quiz_i, whereas goal pursuit persistence is measured by $Card\_Days_i$ and $Quiz\_Days_i$ in this study (Riediger and Freund 2004).

$$\log(Activity_{it} + 1) = \alpha_0 + \alpha_1 \times Adopter_i + \alpha_2 \times After_{it} + \alpha_3 \times (Adopter_i \times After_{it}) + \beta \times D_i + \gamma \times A_{i0} + \tau_t + \varepsilon_{it} \quad (1)$$

The other variables in Model (1) are as follows: $Adopter_i$, *and* $After_{it}$ are dummy variables. $Adopter_i$ denotes whether user *i* adopted mobile. $After_{it}$ denotes the adoption event happened in observational point *t* after user *i* or *i*'s corresponding matched adopter adopted mobile channel. These two variables are the focus of our analyses. In particular, the coefficient of their interaction term indicates the estimated effect of multi-channel adoption. We also control user *i*'s platform usage characteristics D_i (e.g., $Tenure_i$, Cur_Paid_i and $Has\_Playlist_i$), and cumulative goal pursuit activities A_{i0} in the pre-treatment period (i.e., $Num\_Card\_Pre_i$, $Num\_Quiz\_Pre_i$, $Card\_Day\_Pre_i$, $Quiz\_Day\_Pre_i$ and $Num\_Card\_2w_i$,). In the model, τ_t controls for time fixed effect, ε_{it} is the error term. We use the cluster error term at the user level in the estimation. All pre-treatment variables with skewed distribution (i.e., the standard deviation is larger than the mean) are log-transformed. OLS with standard error clustered at the user level was applied in the main DID models in this study.

In the following section, we first examine Hypothesis 1 and explore the underlying mechanism. We also conduct several robustness checks to validate our main results. We then turn our attention to the heterogeneity in goal specificity and goal pursuit

competency to test Hypotheses 2 and 3. Last, we replicate our analyses using another sample.

2.5 Empirical Results

2.5.1 Main Results

We first report mobile adoption's impact on goal pursuit effort and persistence following the Model (1) specification. The results are shown in Table 3. According to the estimated coefficient of the interaction term, it is observed that mobile adopters' goal pursuit effort and persistence increases. On the overall channel, users consume nearly 140.1% more cards and access the platform 0.656 more days after mobile adoption. The increase in goal pursuit effort is also observed in quiz-related learning activities. The number of quizzes and days when a quiz was taken are increased by 111.3% and 0.57, respectively in the post-adoption period. The availability of the mobile channel enables users to expend more effort and perform goal pursuit activities more persistently. In sum, our results indicate that mobile adoption positively affects users' goal pursuit effort and persistence. Therefore, H1a and H1b are supported.

We disentangle mobile adoption's effect on users' goal pursuit by separately estimating the effect on the PC channel. First, with regard to goal pursuit effort (i.e., *Num_Card* and *Num_Quiz*), Table 3 shows the positive effects of mobile adoption are found on PC channel (*Num_Card*: α_3 = 0.900, p<0.001; *Num_Quiz*: α_3 = 0.690, p<0.001). In terms of goal pursuit persistence (i.e., *Card_day* and *Quiz_day*), a similar pattern is observed. Adopters spent 0.387 more days attempting cards and 0.336 more days taking quizzes on their PC after adopting the mobile channel. These results

demonstrate that the positive effect on goal pursuit from mobile adoption also spills over to the traditional PC channel. Users adopting the mobile channel may see an overall increased value proposition from the platform and subsequently invest more effort and time. Importantly, the mobile channel does not substitute the PC channel but rather complements it, as evidenced by the increased goal pursuit effort and persistence across different channels.

Table 3 demonstrates increases in users' goal pursuit effort in both card learning and quiz taking. We further consider the changes in users' effort allocation. We construct a new variable, *Percentage_of_Quiz_Activity*, to reflect the proportion of effort that users spent on quiz related activity, respectively. Note that users may not have any activity in some periods. Estimation without differentiating the zero activity cases would lead to corner solutions (Wooldridge 2002; Burtch et al. 2016). Thus, here our analysis is conditional on the observations wherein users have goal pursuit activities.

Results are presented in Table 4. Note that we only display the results for the percentage of quiz activities as the percentage changes of the card-related activities have the same magnitude as quiz-related results but with the opposite signs. Interestingly, we found that users allocated 4.2% (*PC+Mobile*: $\alpha_3 = 0.042$, $p<0.001$) more of their overall effort in taking quizzes while total effort allocated to card learning decreased. When restricting results to PC-only, we again observe the same direction of changes in effort allocation. More effort is allocated to quiz relevant activities.

Table 3. Main Estimation Results of Users' Goal Pursuit on Overall and PC Channel

	Num_Card		Card_Day		Num_Quiz		Quiz_Day	
	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC
$Mobile_i$	0.023 (0.040)	0.021 (0.038)	0.017 (0.019)	0.016 (0.018)	0.030 (0.035)	0.014 (0.035)	0.024 (0.019)	0.013 (0.019)
$After_{it}$	-0.578*** (0.047)	-0.578*** (0.047)	-0.263*** (0.021)	-0.263*** (0.021)	-0.468*** (0.039)	-0.468*** (0.039)	-0.250*** (0.020)	-0.250*** (0.020)
$Mobile_i \times After_{it}$	1.401*** (0.074)	0.900*** (0.073)	0.656*** (0.033)	0.387*** (0.033)	1.113*** (0.064)	0.690*** (0.062)	0.570*** (0.032)	0.336*** (0.032)
No. of Obs.	7,620	7,620	7,620	7,620	7,620	7,620	7,620	7,620
R-Squared	0.428	0.410	0.434	0.414	0.388	0.367	0.390	0.369

Notes: (1) Clustered standard errors in parentheses: *** $p<0.001$, ** $p<0.01$, * $p<0.05$, + $p<0.1$; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User controls and the time fixed effects are included.

This result can be explained by the characteristics of the mobile channel. Small screen sizes coupled with fragmented and short usage session align better with quiz-taking activities in short sessions rather than longer sessions to learn new cards. Our results indicate that adopting the mobile channel could lead users to develop new habits such as conducting goal pursuit activities they underperformed before.

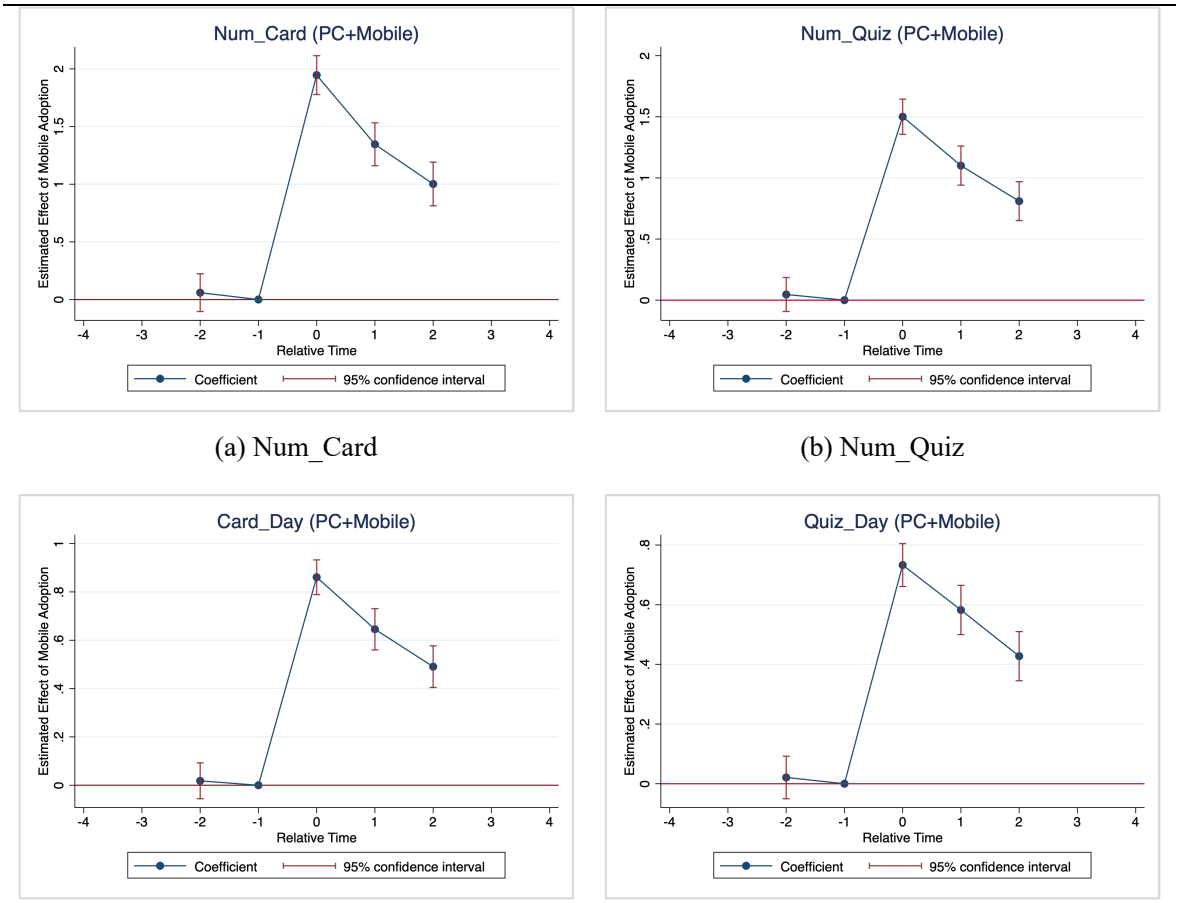
Table 4. Main Estimation Results of Users' Goal Pursuit on the Percentage of Quiz Activities

	PC+Mobile	PC
$Mobile_i$	0.015* (0.008)	0.004 (0.007)
$After_{it}$	-0.026** (0.009)	-0.028** (0.009)
$Mobile_i \times After_{it}$	0.042*** (0.012)	0.025* (0.011)
No. of Obs.	3,986	3,532
R-Squared	0.322	0.373

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) User control and time fixed effect are included.

2.5.2 Using Relative Time Model to Estimate the Impact of Mobile Adoption

Instead of using the standard DID model described in Model (1), we modified Model (1) by replacing $After_{it}$ with relative adoption time to examine the parallel trend assumption of DID; that is, whether individuals significantly change their usage behaviors due to reasons other than mobile adoption that happened prior to the treatment time. Additionally, the results of the modified model could reveal the dynamic impact of mobile adoption.



(a) Num_Card (b) Num_Quiz

(c) Card_Day (d) Quiz_Day

Figure 3. Estimated Effect of Mobile Adoption using Relative Time

The new model specification is shown in Model (2) where T_{nit} is a series of dummy variables representing the relative time of the observation to the week of actual mobile adoption. The period when treatment occurs is coded as 0. *N* in T_{nit} ranges from -2 to +2 to represent the observational window. The observation period leading to the mobile adoption (i.e., designated as time period -1) is utilized as a baseline. Thus, the dummy variable for this observational period is omitted in the estimation (Jung et al. 2019). The rest of the variables remain the same as in Model (1).

$$\log(Activity_{it} + 1) = \alpha_0 + \alpha_1 \times Adopter_i + \alpha_2 \times T_{nit} + \alpha_3 \times (Adopter_i \times T_{nit}) + \beta \times D_i + \gamma \times A_{i0} + \tau_t + \varepsilon_{it} \quad (2)$$

The estimations of Model (2) are shown in Figure 3.[4] No significant effect on the interaction terms is observed in the pre-adoption period. This suggests that adopters' goal pursuit behavior does not change significantly before mobile adoption. Interestingly, once the mobile channel is adopted, significant interaction effects are observed across all three post-adoption periods. Comparing the effect size in three subsequent observational periods after the mobile adoption, we observe that these effects slightly decrease over time. In summary, the results from the model of relative adoption time are consistent with our findings in the main model. We validate the positive effects of mobile adoption on users' goal pursuit effort and persistence.

2.5.3 Robustness Checks

2.5.3.1 Fixed Effect Model

One alternative explanation for the identified effects of mobile adoption is that they are caused by unobserved characteristics of users. To tease out this concern, we utilize a fixed-effect model by taking the impact of user-level time-invariant unobserved factors into account. Model specification of the fixed-effect model is shown as Model (3). In this model, α_i represents the user-level fixed effect. The rest of the variables follow the same representation as variables in Model (1). Results of the fixed-effect model are shown in Table 26 in the Appendix A. The results of the fixed effect model are consistent with the main model.

[4] The complete estimation results is presented in Table 25.

$$\log(Activity_{it} + 1) = \alpha_i + \alpha_2 \times After_{it} + \alpha_3 \times (Adopter_i \times After_{it}) + \alpha_i + \tau_t + \varepsilon_{it} \quad (3)$$

2.5.3.2 Look-ahead PSM (LA-PSM)

Another concern with our main findings is that there may exist endogeneity where some unobserved time-varying factors lead users to adopt the mobile channel and conduct goal pursuit activities. Thus, we further employ look-ahead PSM (LA-PSM) to address this concern (Xu et al. 2016; Kumar et al. 2018; Bapna et al. 2018). With LA-PSM, it is assumed that all adopters share similar observed and unobserved characteristics, such as intrinsic motivation. Thus, instead of finding a control group from non-adopters as commonly done in traditional PSM, we find the control group from future adopters with LA-PSM. That is, we match adopters in the scrutinized observation windows with users who eventually adopt mobile in the future to control the effect of potential unobserved time-varying factors.

To operationalize LA-PSM, for each adopter, we find a matched user from those who adopt the mobile channel at least eight weeks after the focal adopter. Aside from this difference, the rest of the matching procedures are the same as the main model. In total, 389 matches are generated using LA-PSM. With the newly matched samples, we perform the DID model again and present the results in Table 27 in the Appendix A. Overall, the estimation results using LA-PSM are consistent with results obtained from the main model, which further demonstrates that our conclusions are robust.

2.5.3.3 Alternative Matching Method

To further validate our results, another matching strategy, coarsened exact matching (CEM), is performed. Compared with PSM, CEM uses stricter criteria to eliminate the imbalance in the original dataset (Lacus et al. 2009). We apply CEM with k2k to generate 156 matched pairs and then re-estimate Model (1). Results are presented in Table 28, and they are consistent with PSM-based results.

2.5.3.4 Falsification Test

As another robustness check, we performed a falsification test by selecting another week from the pre-adoption period as the artificial adoption week and then once again estimating the effect of mobile adoption. If there are no spurious events occurring in the actual pre-adoption period that induces change in users' goal pursuit behavior, there should be no observable significant effect from the assigned mobile adoption week.

To operationalize this falsification test, we follow Jung et al.'s (2019) approach and narrow down our data to only observations in the pre-adoption phase. Then the middle time of that period (i.e., a week before the actual adoption) is used as the hypothetical adoption week for each user. Results are presented in Table 29, and no significant effect is observed. It suggests that users' goal pursuit behaviors did not change before their mobile adoption. The results indicate that the positive effects we observed in the main models are driven by mobile adoption but not due to other spurious factors.

2.6 Heterogeneity Test

2.6.1 Model Specification

We further conduct a series of heterogeneity tests to examine hypotheses H2-H3 using Model (4). Compared to the main model, we add one more interaction term, $Adopter_i \times After_{it} \times Heterogeneity_i$, to Model (4) where $Heterogeneity_i$ represents user i's goal specificity and goal pursuit competency.

$$\log(Activity_{it} + 1) = \alpha_{0t} + \alpha_1 \times Adopter_i + \alpha_2 \times After_{it} + \alpha_3 \times (Adopter_i \times After_{it}) + \alpha_4 \times (Adopter_i \times After_{it} \times Heterogeneity_i) + \alpha_5 \times (Adopter_i \times Heterogeneity_i) + \alpha_6 \times (After_{it} \times Heterogeneity_i) + \beta \times D_i + \gamma \times A_{i0} + \tau_t + \varepsilon_{it} \quad (4)$$

2.6.2 Goal Specificity

Besides offering video-based lessons to assist users in learning medical knowledge, Picmonic also provides a unique playlist by organizing the United States Medical Licensing Examination (USMLE[5]) related content to help medical students better prepare for the exam accordingly. In particular, to help users learn and practice USMLE relevant content, Picmonic designs USMLE-focused pathways (i.e., *USMLE Step 1* and *Step 2*) by organizing content in a manner that aligns with USMLE format. Additionally, users can access all learning content through two other pathways, *Courses* and *Body Systems*, in which the content is organized with a more generic learning purpose. Given the content organization logic behind each pathway, we utilize the pathway usage as a proxy for users' goal specificity. Users who have accessed content through the USMLE-focused pathway are considered to have a more specific goal and stronger motivation when using

[5] A standardized test that medical students must pass to be eligible for a medical license.

the focal platform. Conversely, users who never accessed USMLE-related pathways are categorized with relatively less goal specificity. We examine the moderating role of goal specificity by comparing mobile adoption's impact on goal pursuit between these two groups of users.

Table 5. Heterogeneity Test on Goal Specificity

	Goal Pursuit Effort		Goal Pursuit Persistence	
	Num_Card	Num_Quiz	Card_Day	Quiz_Day
$Mobile_i$	0.067 (0.048)	0.039 (0.043)	0.035 (0.024)	0.033 (0.024)
$After_{it}$	-0.520*** (0.054)	-0.415*** (0.045)	-0.243*** (0.023)	-0.221*** (0.023)
$Mobile_i \times After_{it}$	1.143*** (0.092)	0.904*** (0.079)	0.547*** (0.041)	0.464*** (0.040)
$Mobile_i \times USMLE\_focused\_users_i$	-0.200* (0.086)	-0.120 (0.078)	-0.081+ (0.042)	-0.071+ (0.042)
$After_{it} \times USMLE\_focused\_users_i$	-0.178+ (0.103)	-0.164+ (0.087)	-0.065 (0.046)	-0.090* (0.045)
$Mobile_i \times After_{it} \times USMLE\_focused\_users_i$	0.565*** (0.154)	0.466*** (0.133)	0.235*** (0.069)	0.238*** (0.068)
No. of Obs.	7,620	7,620	7,620	7,620
R-Squared	0.437	0.398	0.442	0.399

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User control and time fixed effect are included.

Table 5 shows the results of the heterogeneity test in terms of goal specificity. We found that compared to users with generic learning purposes, USMLE-focused users exert 56.5% more effort and persist 0.235 more days in cards-related goal pursuit activities after mobile adoption. The same effects are also observed in quiz related goal pursuit activities. These results suggest that users with a more specific goal (i.e., passing

a standardized exam) achieved a higher level of improvement in their goal pursuit from adopting the mobile channel, which aligns with what we expected in the hypothesis. Overall, H2 is supported.

2.6.3 Goal Pursuit Competency

The principal value proposition of our focal platform is their creative medical content to help users learn visually. These contents correspond to different stages of medical school study starting with more foundational courses and proceeding to more advanced and specialized ones. Users taking more advanced courses are likely to have higher competency as reflected by a deeper understanding of medical-related content and more experience in goal pursuit. Thus, within the context of this study, the course progress exhibited by users is reflective of their goal pursuit competency. We perform a heterogeneity test in users' goal pursuit competency in this section.

Operationally, we categorize medical content on Picmonic as foundational or specialized according to the typical medical school curriculum[6]. Individuals who have attempted specialized courses are identified as high goal competency users. Otherwise, they are categorized as low goal competency users. We perform the heterogeneity test regarding users' goal pursuit competency, and the results are presented in Table 6. From this table, we observe high competency users did not perform statistically differently from users with low competency in terms of card-related goal pursuit effort. However, some differences are found between these two types of users in quiz-related activities.

[6] Fundamental courses include pathology, pharmacology, biochemistry, microbiology, physiology, behavior & psychiatry, anatomy & embryology, and reproductive system. Specialized courses include dermatology, surgery, pediatrics, obstetrics & gynecology, and internal medicine.

We found that high goal competency users exert 29% more effort and persist 0.142 more days in attempting quizzes than users with low goal competency level. These results demonstrate that the heterogeneity in users' goal pursuit competency might be subtle and contingent on the goal pursuit activities. Our empirical results partially support H3.

Table 6. Heterogeneity Test on Goal Pursuit Competency

	Goal Pursuit Effort		Goal Pursuit Persistence	
	Num_Card	Num_Quiz	Card_Day	Quiz_Day
$Mobile_i$	0.004 (0.044)	0.011 (0.039)	0.012 (0.021)	0.014 (0.021)
$After_{it}$	-0.582*** (0.051)	-0.462*** (0.042)	-0.263*** (0.022)	-0.242*** (0.022)
$Mobile_i \times After_{it}$	1.314*** (0.083)	1.021*** (0.070)	0.622*** (0.038)	0.529*** (0.036)
$Mobile_i \times$ Specialized-focused_$User_i$	-0.146 (0.110)	-0.128 (0.103)	-0.082 (0.055)	-0.056 (0.054)
$After_{it} \times$ Specialized-focused_$User_i$	0.021 (0.127)	-0.039 (0.112)	1.33e-05 (0.058)	-0.045 (0.057)
$Mobile_i \times After_{it} \times$ Specialized-focused_$User_i$	0.248 (0.180)	0.290+ (0.160)	0.103 (0.082)	0.142+ (0.081)
No. of Obs.	7,620	7,620	7,620	7,620
R-Squared	0.440	0.401	0.445	0.399

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User control and time fixed effect are included.

2.7 Discussion

This research utilizes a quasi-natural experiment approach to examine multi-channel adoption's impact on users' goal pursuit. Our results suggest that users' goal pursuit

effort and persistence are enhanced after adopting the mobile channel. Interestingly, we observe the enlarging gap between different types of students in our study. Users with a more specific goal and higher goal pursuit competence benefit most from utilizing the additional goal-pursuit means. They achieved a more considerable improvement in their goal-pursuit effort and persistence.

2.7.1 Theoretical Implications

This work contributes to the nascent literature on technology-mediated goal pursuit (Uetake and Yang 2018). Prior goal pursuit literature mainly emphasizes goal pursuit behaviors without considering the role of technology (Fishbach and Finklstein 2012). In contrast, this research focuses on technology adoption and examines its impact on users' goal pursuit. Our results highlight additional channel significantly improve users' goal pursuit effort and persistence. Additionally, given the affordance and constraints of the mobile channel, users' goal pursuit activities become more ubiquitous and continuous after mobile adoption. We also found that users adjust their usage pattern to test their knowledge instead of passively watching content.

Moreover, this study also contributes to multi-channel adoption literature (Xu et al. 2017; Liu et al. 2016) by providing a new theoretical angle (i.e., goal pursuit theory) to study the impact of multi-channel adoption. Based on the unique empirical observations from the setting of goal-directed platforms, in addition to examining the positive impact of mobile adoption on users' subsequent activities on platforms, we further dig into users' goal pursuit characteristics and demonstrate the heterogeneity in users' goal specificity and goal pursuit competency. Our results indicate that users benefit

from adopting the additional channel differently. In particular, users with a more specific goal and higher competency achieved higher improvement levels in the post-adoption period.

2.7.2 Practical Implications

From a practical perspective, this study indicates that platform owners can facilitate users' goal pursuit by investing in mobile channel development. The affordance introduced by this new channel will enable users to diversify their goal pursuit activities and utilize the platform's resources more thoroughly. Further, users may develop more intensive and frequent platform usage habits. Such changes will result in more frequent and persistent effort in goal pursuit.

When investigating resources to the mobile channel, it is also vital for platform owners to maintain the traditional PC-based channel to leverage complementarity between these two channels. Our results indicate that as users exert more effort and persistence on the newly introduced mobile channel, this new usage pattern also migrates to the PC-based channel.

Platform owners need to be aware of the heterogeneous impact of multi-channel adoption on users with different goal pursuit characteristics. On the one hand, platforms can strategically utilize these heterogeneities to better target persuasive messages to individual users to increase user engagement. For example, in the early stages of a new channel's release, platform owners could invite users with more specific goals to adopt the channel as these users are more likely to achieve significant improvement in their goal pursuit. On the other hand, platform operators should also offer more help to users

who benefit less from mobile channel adoption and then help them achieve more improvements in their goal pursuit. Potential interventions include offering usage training, reminding users to set up specific goals on their learning, and providing more feedback on channel usage. In sum, understanding the heterogeneous effects of mobile adoption can help platforms improve user engagement and ultimately achieve better business performance.

2.7.3 Limitations

We note some limitations in this work. It would be interesting to go beyond the goal pursuit effort and persistence to examine further the impact of multi-channel adoption on users' goal pursuit performance. However, given the business focus of the studied platform, relevant goal pursuit performance data (i.e., finals scores on USMLE exams) is not available for us. Despite that, it is a promising avenue for future work in technique-augmented goal pursuit literature.

Goal pursuit is a systematic process, and the impact of new goal pursuit methods might be moderated by numerous factors such as users' perception and goal pursuit stage (Huang et al. 2011). One of the weaknesses of using an observational dataset is that we cannot observe users' perceptions and intentions. It would be helpful to integrate our current archived data-based analyses with perception or intention-focused lab experiments. Observations from the latter will complement our current empirical findings and contribute to a deeper understanding of new channel adoption in goal-directed platforms.

Similar to other context-dependent studies, our conclusions may have generalizability concerns considering that the samples all come from medical students. We try to address this limitation by replicating our analyses using nursing students, the second biggest group on Picmonic. Nursing students can be considered a distinct user group compared to medical students due to demographic differences (e.g., age, educational background) and instructional program. For example, in the United States, nursing students are generally undergraduates, whereas medical students are typically graduate students. Such differences between these two groups of people could result in their different attitudes towards technology adoption and usage. If the results from the sample of nursing students are consistent with the estimations with samples comprised of medical students, it would be good supplementary evidence to our previous findings and demonstrate the generalizability of the effects of multi-channel adoption on goal pursuit.

We follow the same identification strategy and data analyses as applied to medical students. Results are displayed in Appendix B. Overall, results when utilizing the sample of nursing students are consistent with our previous experiment. Interestingly, in terms of goal pursuit effort (i.e., *Num_Card* and *Num_Quiz*), that complementarity between the PC channel and the mobile channel for nursing students only persists for a relatively short period of time. One possible explanation is that the age and gender composition of medical and nursing students are different. Thus, nursing students are more likely to favor the mobile over the PC-based channel.

CHAPTER 3

THE EFFECTS OF MACHINE-POWERED PLATFORM GOVERNANCE: AN EMPIRICAL STUDY OF CONTENT MODERATION

3.1 Introduction

Dominant online platforms have been facing safety challenges and economic loss caused by inappropriate content such as hate speech and trolling (Matias 2019 a; Roberts 2014). Diverse and complex online environments increase the need for devising platform policies, and consequently content moderation when content policies are violated. Given the need for monitoring online content, platforms heavily rely on human labor to maintain their content. For example, Facebook hires professional teams and incorporates user reports to detect harmful content (Birman 2018; Catherine et al. 2016; Menking and Erickson 2015). Reddit and Wikipedia (Zheng et al. 2019) rely on volunteer moderators to design the community rules and authorize these human moderators to manage their community members and content (Matias 2019 a; Jhaver et al. 2019 b, 2017).

Although platforms may adopt different approaches to organizing their moderation team, volunteer moderators have become an important workforce for governance in digital platforms. Platforms such as Reddit and Wikipedia rely entirely on uncompensated labor to sustain their businesses. However, due to the complex online environment and the increasing demand for content moderation, volunteer moderators usually experience burnout while attempting to maintain a healthy and thriving online community (Dosono and Semaan 2019; Grimmelmann 2015; Seering et al. 2017). Meanwhile, this unpaid labor is usually expected to balance unfulfilled expectations from

the community and handle the dark side of the internet, such as abusive content and harassment (Gillespie 2018; Matias 2019 b). These potential threats raise a rising concern about the stability and sustainability of volunteer-based governance.

Algorithm-based moderation opens a new avenue for platform governance (Gollatz et al. 2018; Hammer 2016; Seering et al. 2019). For instance, Facebook has applied advanced machine learning techniques to detect pornographic content before it can be viewed and shared (Robert 2014; Gillespie 2018). On Reddit, many communities use bot moderators to screen participants and submissions, send reminders, and remove inappropriate content (Chandrasekharan et al. 2018). By embedding moderation rules and processing logic into algorithms, platforms can automate simple and routine moderation tasks. Compared to humans, algorithm-based moderation has high scalability to handle massive online content comprehensively and instantly.

The usage of machine-assisted moderation has attracted growing attention in recent years. Prior literature suggests that applying algorithms can effectively offload volunteer moderators and enforce community norms (Jhaver et al. 2019a). While platforms are moving forward to the technological mode of governance, some researchers pose questions regarding de-humanization, given the critical role of interaction between moderators and users in sustaining the platform (Ruckenstein and Turunen 2019; Yu et al. 2020; Karusala et al. 2017). De-humanization concern comes from two aspects. First, volunteer moderators are both community managers and community users. Given the dual role, when moderation tasks get automated, human moderators may walk away from their community manager role and decrease their interaction with community users. Second, machines can also make volunteer moderators feel empowered and encourage

them to perform more policing work while reducing their efforts at more supportive types of governance.

Delegating moderation tasks to machines can, however, also result in an increase in volunteers' engagement. The automation of moderation tasks will offer a more efficient and healthier working environment for human moderators. The improved working environment can in turn enhance their commitment and engagement to their moderation tasks (Alfes et al. 2016; Smith 1994). Simultaneously, volunteers can also go beyond enforcing compliance in community rules by devoting more effort to advanced moderations that require more care (Karusala et al. 2017). Despite the increasing adoption of algorithm-based governance by digital platforms, relevant academic research is still in its infancy. A few prior studies have investigated the human-machine collaboration in platform regulation (Ren and Kraut 2014; Luo et al. 2019). Several studies focusing on bot usage are mainly from qualitative perspectives with data from a few selected online communities (Jhaver et al. 2019 a; Ruckenstein and Turunen 2019). To fill this research gap, we investigate the impact of algorithm-based moderation tools (herein referred to as 'bot moderators') on human moderators' behavior and examine whether the technological mode of regulation leads to de-humanization. Formally, we want to approach this research objective by investigating two research questions:

RQ1: How does introducing machine-powered platform governance (i.e., bot moderators) affect human moderators' moderation-related effort?

RQ2: How does introducing bot moderators affect the amount of effort human moderators spent on different types of moderation (i.e., corrective and supportive moderation)?

To answer our research questions, we select Reddit as our research context, given its highly autonomous communities and well-documented moderation records. The Reddit platform has thousands of communities (i.e., subreddits), with only a small set of volunteers. With the rapid increase of user base on Reddit, several bot moderators have been created to facilitate routine content moderation. When a bot moderator identifies a rule violation on Reddit, it will follow the community guidelines and take actions such as removing content or banning users. Meanwhile, bot moderators leave a moderation record to inform affected users in the form of public comment. This design enables us to observe the adoption of bot moderators on each subreddit and examine their impact on human moderators' behaviors in respective subreddits.

We collect moderation records and user and human moderator participation from 156 subreddits on Reddit from 2013 to 2014. We identify the automation of moderation tasks from the public moderation records by bot moderators. We employ an advanced natural language processing technique, BERT (Bidirectional Encoder Representations from Transformers), to identify human moderators' activities. A Difference-in-Differences model is then applied to estimate the impact of moderation automation. Results from our econometric analyses suggest that after bot moderators are implemented, volunteer human moderators perform more moderation-related activities. Human moderators make 20.2% more corrective moderations to enforce community guidelines. Meanwhile, they also provide their community with more supportive

comments — offering 14.9% more explanations. Notably, the effect manifests primarily among communities with large user bases and detailed guidelines, suggesting that community needs for moderation are the key factors driving more voluntary contributions. Overall, our results indicate that delegating content moderation tasks to algorithms augments volunteer moderators' role as community managers. The increased moderation-related engagements, especially supportive behaviors, alleviates the common concerns of de-humanization in the automation of platform governance.

This research contributes to three different streams of literature. We first contribute to the platform governance literature for digital platforms and online communities (Ren and Kraut 2014). Prior literature mostly concentrates on governance facilitated entirely by human labor. Our study contributes to this stream of work by examining the new governance mode with humans' and machines' collaborative effort. Second, our work also contributes to the human-machine frontier. Human-machine collaboration has recently been studied in contexts such as online commerce (Luo et al. 2019; Schanke et al. 2021; Bai et al. 2020). Our research specifically focuses on the impacts of algorithm-based moderation tools on volunteer content moderation. Our empirical results indicate that bot moderators complement rather than replace human moderation in achieving more comprehensive moderation. Last but not least, our study contributes to the field of computer-supported cooperative work by providing one of the first large-scale empirical evidence for the emerging discussion about the de-humanizing effects of content moderation (Zheng et al. 2019; Dosono and Semaan 2019; Matias 2019b). We found that applying algorithms to moderate online content will not drive volunteer human moderators to reduce their effort in their moderation roles. In contrast,

algorithm-based moderation tools empower and stimulate voluntary human moderators to achieve more moderation tasks. Moreover, we see more supportive and caring engagements from human moderators after tedious and routine tasks get automated by bot moderators.

3.2 Related Literature & Hypotheses

3.2.1 Platform Governance and Human-machine Collaboration

Platform governance is the mechanism that regulates individuals' participation in a community to increase the interaction quantity, quality and prevent abuse (Grimmelmann et al. 2015). Given the rapid growth of online participants and increasingly diverse content, maintaining platforms is a major challenge for platform operators. Most platforms rely on human labor to manage communities and moderate massive amounts of online content. Meanwhile, those human moderators face extensive physical and mental pressure from the large scale of maintenance work and user misconduct, such as harassment and trolling, to even more severe violations.

With the increasing use of algorithms in digital platforms, there is a growing interest in human-machine collaboration in platform governance. Machines can be used at different stages of regulation (i.e., before or after the rule violation happens). Matias (2019a) studies bot-generated reminders, a proactive type of moderation, and its impact on discussion participation. He conducted a randomized experiment on a science discussion community on Reddit. He found that presenting a sticky announcement with community rules to discussion threads will encourage more newcomers to participate in

discussion. Moreover, unruly and harassing conversations decreased with the presence of highly visible reminders.

A large body of research focuses on reactive regulation (Srinivasan et al. 2019; Jaidka et al. 2019; Seering et al. 2017). For example, Srinivasan et al. (2019) investigate the role of comment deletion in users' future engagement, achievement, compliance, and content toxicity. Results suggest that removing problematic comments will lead to an immediate decrease in noncompliance rates. Jhaver et al. (2019a) interviewed volunteer moderators on Reddit and documented the benefits and challenges of bot-assisted moderation. On the one hand, bots can effectively identify hate speech, personal attacks, and other inappropriate content. On the other hand, bots may cause false-positive moderations and over-policing. Therefore, moderators need to acquire bot-related technical skills to regularly correct false positive moderations, adjust bot functions along with the community rule changes, and maintain an engaging community environment.

Overall, there has been increasing attention paid to the human-machine collaboration in platform governance. However, prior work provides limited quantitative empirical evidence on bot moderators' impact on human moderators' voluntary engagement.

3.2.2 Volunteer Human Moderators and Moderation-related Activity

The influence of machines on human labor has been discussed in the literature in recent years (Dixon et al. 2020; Luo et al. 2019). For example, Dixon et al. (2020) study the usage of robots in the manufacturing sector. They found that as companies utilized more robots in production, the overall employment increases, but employment for manager

positions decreases. In addition to the traditional business sector, machines and algorithms are also widely applied to online businesses. Luo et al. (2019) examine the functionality of chatbots in outbound sales calls. They found that when the identity of chatbots is not disclosed, their performance will be similar to proficient workers and even four times higher than inexperienced workers.

Overall, prior studies come mainly from a functional perspective to investigate the relationship between humans and machines and their collaborative outcomes. By comparing their work capabilities, machines have shown the potential to substitute human in positions that emphasize routine and standard duties (Autor and Dorn 2013). From this perspective, in the context of online content moderation, machines and algorithms can also substitute humans in achieving large-scale moderation tasks. Specifically, Jhaver et al. (2019 c) study the perceived difference between bots and human moderators on Reddit. Their interview results suggest that as long as platform moderations maintain a high level of transparency, there is no difference between the removals executed by human moderators and those by bot moderators. Both approaches are effective at educating users and reducing the policy violation of future submissions. This functional substitution can result in a decreased need for human moderators to engage in community governance. Thus, human moderators could be less engaged in their regulator role when communities adopt bot moderators for platform governance.

However, not all moderation tasks can be entirely handled by bot moderators. Compared to humans, bot moderators have limited capability to understand contexts and subjects. Some high-level community rules such as detecting satirical languages and hate speech still rely on and are subject to human interpretation. Moreover, functional

substitution is not the only driver of changes in human moderators' behavior. Human moderators, particularly those moderating online communities, are recognized as different from the compensated workforce in prior literature as they are voluntary labor on the platform. Their moderation-related engagement is driven by their internal motivation and content consumption needs. Abundant literature on volunteerism suggests that task-related and emotion-related support are two key factors in volunteers' engagement and commitment (Alfes et al. 2016; Smith 1994; Bang et al. 2013 b; Vecina et al. 2013). Automating routine moderation tasks with bot moderators will offer human moderators a more supportive environment. First, machines can offload human moderators from routine and tedious moderation tasks (e.g., posts that do not meet minimal length requirement) and make their working environment more productive and efficient. Prior literature also indicates that enhanced working efficiency in the non-profit sector leads to increased volunteer engagement and commitment (Smith 1994). Machines will also empower human moderators; such that human moderators will be less likely to suffer from burnout and thus can better accommodate the other aspects of content moderation needs in larger quantities for their respective communities.

Bots can also provide human moderators with more emotional support, thus creating a more pleasant working environment. In addition to dealing with a large number of moderation tasks, human moderators also suffer from emotional stress in the diverse and complex online environment (Matias 2019a). They usually need to handle potentially harmful content and uncivil submissions. Bots can help human moderators filter out inappropriate content first and speed up the content moderation process. Delegating certain moderation tasks to bot moderators can therefore protect moderators from

harassment and harmful content (Jhaver et al. 2017). A healthy and safe working environment will make volunteer work more attractive and stimulate more voluntary engagements from human moderators.

Additionally, bots can strengthen the human moderators' role as community managers. Compared to the role of a regular community member, the human moderator role brings volunteers status and prestige in the community (Yang et al. 2019; Anzalone 2020). This reputation can also serve as incentives for volunteers to maintain their effort and exposure in the community. Although certain types of their moderation tasks and expected workload will decrease after bot moderators are implemented in their community, they will likely continue their behaviors as community moderators but not stopping contributing.

Considering the enhanced work environment and the empowerment effect from bot moderators, we propose our first hypothesis:

H1: Delegating moderation to bots will lead human moderators to engage in more moderation activities.

3.2.3 Human Moderation Types

In addition to human moderators' overall contribution to community moderation, it is essential to uncover deeper insights into how automating content moderation to bot moderators would affect the moderation focus. By interviewing 56 volunteer moderators from three major online content platforms (i.e., Reddit, Twitch, and Facebook), Seering et al. (2019) documented a list of general duties that moderators need to fulfill. Based on governance priority and effort that a moderation requires, these duties can be classified

into two categories: basic and advanced moderations. Basic moderations, such as monitoring community activity and removing rule-violating content, are the top priority for human moderators because these efforts are critical to the safety and integrity of a community. In contrast, advanced moderations, such as responding to community members, seeding content, and contributing to discussions, are corresponding effort to foster a more engaging community. It requires moderators to invest more effort into the community nurturing in addition to enforcing the rule compliance.

Similarly, Ruckenstein and Turunen (2019) summarize two major types of content moderations: logic of choice and logic of care. The logic of choice is a *corrective* type of governance (i.e., removing inappropriate content). In contrast, the logic of care emphasizes *supportive* moderation (i.e., explaining community rules and offering suggestions to community members). Both corrective and supportive moderation are critical factors for community building and growth. On the one hand, *corrective* moderation is an essential component for communities to maintain their integrity and order. Srinivasan et al. (2019) studied a typical type of corrective moderation—content removal—and found that deleting problematic content would reduce community noncompliance rates. *Supportive* moderation, on the other hand, would stimulate healthier and more constructive interactions in the community. It is associated with more desired content for their community members and works as the driving force for community sustainability (Karusala et al. 2017; Yu et al. 2018). Supportive moderation complements corrective moderation in achieving more effective platform governance. For example, Jhaver et al. (2019 c) found that while executing content removal, offering sufficient explanations would significantly reduce the possibility of users' future rule

violations. In recent years, researchers have also begun to call for more care during content moderation (Yu et al. 2018; Yu et al. 2020) while corrective moderations become easier to implement after bot moderators are widely implemented into platform governance.

Given the importance of both corrective and supportive moderations, it is critical to investigate how bots affect either type of human moderation, especially the supportive moderation that is more desired. Theoretically, as platforms automate moderation tasks, volunteers' moderation focus may change. First, the limitation of bot moderators suggests that human intelligence is difficult to completely substitute with algorithm-based tools, especially in scenarios that require understanding context and handling unusual situations (Jhaver et al. 2019 a; Seering et al. 2019). When bots become available, human moderators can be relieved from burdensome moderation tasks. Thus, they will have more bandwidth to enforce guidelines that are difficult to adjudicate (Jhaver et al. 2019 a). Therefore, corrective/policing behavior may increase.

Moreover, when moderators are free from tedious and routine moderation tasks, they have more capacity to go beyond basic policing tasks and perform more advanced moderation work such as supportive activities. Butler et al. (2002) found that compared to regular users, community members with former leadership roles are more likely to contribute to community building such as helping community members. Considering human moderators are community managers, they are likely to engage with more supportive moderation behaviors when they are empowered by the bot moderators. Therefore, we propose our second hypothesis:

H2: Applying bots to content moderation will encourage human moderators to generate more corrective and supportive engagements.

3.3 Empirical Setting

3.3.1 Research Context

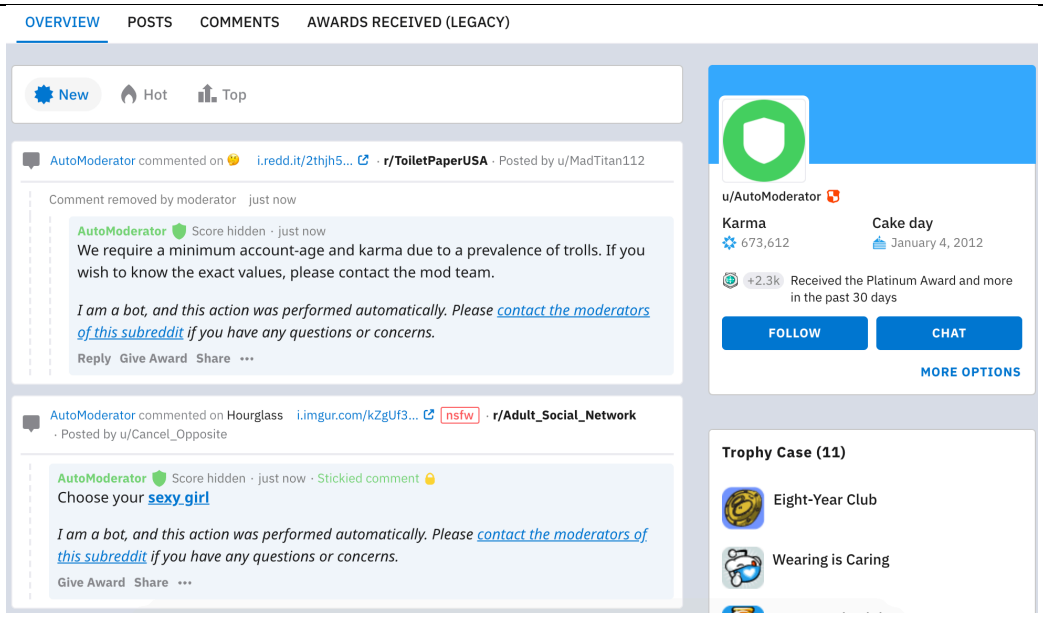
We collected our data from Reddit, a large online discussion and community website. It comprises more than 130,000 active communities, termed subreddits, covering a wide range of subjects such as world news, sports, writing, and movies. Like other dominant social media platforms such as Twitter, users can post content, leave comments, and upvote/downvote others' content on Reddit. Currently, Reddit has more than 430 million monthly active users, and it has become the fifth most visited user-generated content platform in the world as of December 2019.[7]

Reddit used to solely rely on volunteers to manage all their communities. A majority of human moderators are community users who have a solid understanding of the community identity and rules. These volunteer human moderators design community rules and monitor community activity regularly. When a rule violation happens, they are also authorized to remove content and suspend users.

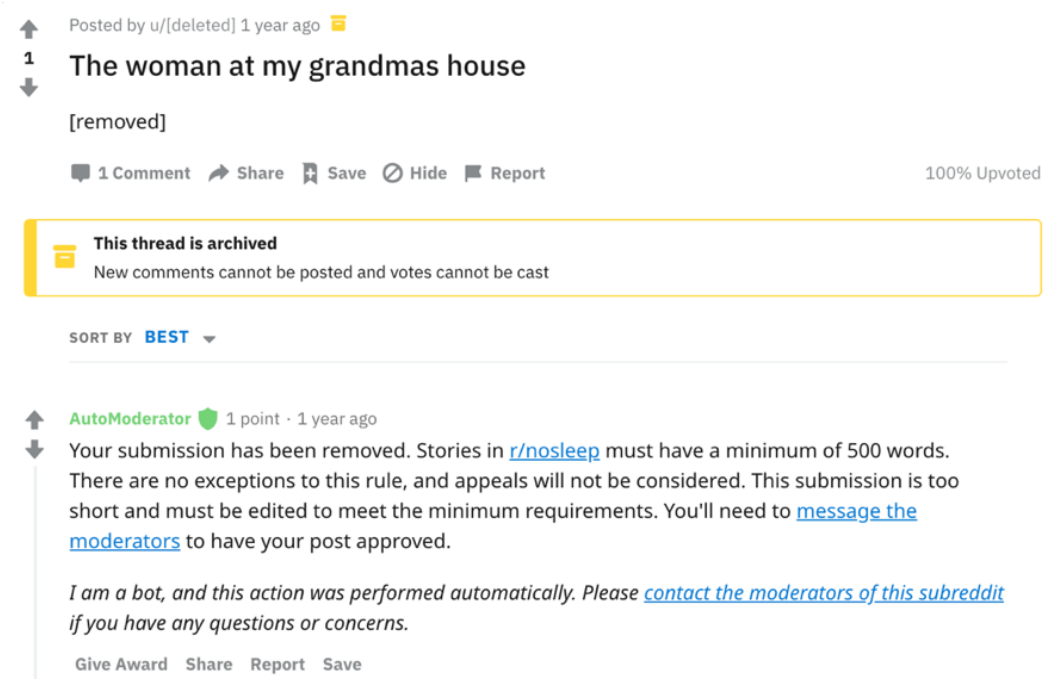
Since 2012, subreddits started to utilize bot moderators to assist their platform governance. Two characteristics make Reddit to be an ideal research context for our study: data availability and transparency. Due to security and intellectual property concerns, most platforms do not disclose details of the technical strategies used for their

[7] https://redditblog.com/2019/12/04/reddits-2019-year-in-review/

content moderation. Moreover, for those platforms, we neither observe moderation records for human moderators nor bot moderators. In contrast, Reddit makes both task delegation and moderation records available to the public. On Reddit, subreddits use bots to assist with their community governance (Chandrasekharan et al. 2018; Chandrasekharan et al. 2019; Fiesler et al. 2018; Long et al. 2017). These bot moderators can automatically inspect newly submitted content and conduct a series of actions accordingly. Each bot moderator has its profile page, which lists its creation date and historical moderation records. Most importantly, all records are presented in the form of public comments, as shown in Figure 4. These comments document details such as moderation time, action, explanations, and suggestions to the content creator. Human moderators perform their moderations in a similar manner. When they find a rule violation, they can perform moderation and inform the content creators through public comments, submission flairs, and private messages. Among these three communication channels, public comment is the dominant approach (Jhaver et al. 2009 b). Public moderation records are the primary data source for our study. And based on these records, we recover the moderation automation timeline and investigate its impacts on human moderators' voluntary activities.



(a) Profile page of AutoModerator



(b) The Screenshot of a Moderation Record by AutoModerator

Figure 4. AutoModerator and Its Moderation Records

3.3.2 Data Collection and Measures

We choose 'AutoModerator' as the focal bot in our study. AutoModerator was introduced in 2012, and it has become the most influential bot moderator on Reddit. More than 4,000 subreddits have adopted AutoModerator as of December of 2019. The bot generates thousands of moderation records every day. AutoModerator is a moderation tool with various functions. To delegate tasks to AutoModerator, human moderators need only follow simple syntax and add bot-related codes into their community meta page. The typical moderation tasks that AutoModerator can achieve include restricting users under a certain level of reputation on the platform and preventing users from posting content from sources that are blacklisted. Subreddits on the Reddit platform can integrate the AutoModerator functions to deploy it as a bot moderator based on their community guidelines and moderation needs.

Our data collection includes four steps. First, we collect all moderation records by AutoModerator from 2013 to 2014 using PushShift API.[8] We choose this observation period because AutoModerator's function was relatively stable after being released for a year. Further, more subreddits started incorporating it into their moderation since 2013. Thus, the selection of subreddits in that period would contain ample variation in the corresponding empirical analyses. Second, we use the time when the first AutoModerator's moderation was performed as the proxy adoption date of AutoModerator in a subreddit. With the complete moderation records collected in the first step, we observe that 156 subreddits adopted AutoModerator from 2013 to 2014. We

[8] https://pushshift.io/

report the complete list of subreddits in Appendix C. Next, we identify human moderators for each of the studied subreddits. Reddit grants moderators a "mod" flair to signal their role as community managers, and this flair is shown along with moderators' comments. Taking advantage of this platform design feature, we obtain the list of users with this flair and further collect their comments during the observation window. Lastly, with AutoModerator's and human moderators' comments, we perform machine learning to extract the automated rules and human moderators' engagement types.

We performed two aspects of data pre-processing. First, on the AutoModerator side, we extract automated moderation tasks and task implementation time to recover the timeline of governance delegation. We utilize this timeline to construct the independent variable of interests (i.e., $\text{AutoMatedTasks}_{it}$ and After_{it}) in our following analyses. Given that AutoModerator mostly follows limited formats to frame a moderation record, to achieve this goal, we can employ a rule-based approach to extract the automated task and its implementation time.

Second, on the human moderator side, our pre-processing task is to identify their activities and moderation types. We seek to perform a theory-driven annotation (labeling) task for a sample of human moderators' comments for four categories: policing, explanation, suggestions, and casual talk. Table 7 illustrates examples for each category. Among these four categories, policing, explanation, and suggestion belong to the moderation-related comments, whereas casual talk represents the informal interaction that human moderators have with community users. With regards to the moderation type, policing represents corrective moderation, whereas suggestions and explanations are

supportive moderations.[9] Compared to processing of bot comments, this task is relatively challenging due to the diversity and complexity of human language. Our methods need to go beyond the vocabularies to grasp the semantic meaning and purpose of a comment.

Table 7. Examples of Different Types of Moderator Comments

Category	Example
Policing	"Sorry, this is a repost within 3 months. Therefore, I am removing it."
Explanation	"This has been removed because you did not include the resolution in the title and because of man-made structures and you did not include the location in the title."
Suggestion	"There are instructions in the FAQ as well as our other submission guidelines. This would be more appropriate in r/villageporn. Feel free to resubmit with the resolution in the title. Thanks!"
Casual talk	"Thank you and it is even more beautiful in the reality especially the very special light of Iceland"

Specifically, we utilize BERT, a deep learning approach, to classify human moderators' moderation types (Devlin et al. 2018). BERT was proposed in 2018, and it has quickly become the state-of-the-art natural language processing (NLP) technics and has achieved exceptional performance in NLP tasks such as classification, Q&A, and commonsense reasoning. Compared to traditional NLP methods, BERT is a bidirectional language model trained on a large-scale corpus. It has a better sense of the language context and the relationship between all words regardless of their respective position. Therefore, BERT can capture the meaning of comments regardless of the vocabulary

[9] Following real-world practice, a comment can only be either moderation-related or not, but a moderation-related comment can have both corrective and supportive components. Thus, in our annotation task, we restrict casual talk as exclusive from three moderation-related categories, and we allow the three moderation types to coexist in a human moderator's comment.

choice and sentence sequence. We use BERT-based classifiers to identify moderation-related comments as well as detailed moderation types.

Table 8. Performance of BERT-based Classifiers

Category	Accuracy	Precision	Recall	F1 Score
Policing	97.34%	96.99%	95.68%	96.33%
Explanation	95.37%	95.02%	95.22%	95.12%
Suggestion	95.96%	96.84%	95.15%	95.99%
Casual talk	92.52%	94.47%	84.66%	89.30%

Operationally, similar to other deep learning-based NLP techniques, we use the pre-trained BERT model and then fine-tune the parameters with a set of labeled data. Pre-trained BERT can capture the context-free meaning of the comments. The labeled data will provide more details about our context. We obtain the pre-trained model from Google Research, and this model is trained on large corpora collected from Wikipedia and book chapters.[10] Then, we randomly sampled 1,017 comments from our whole dataset and manually labeled these comments for each category. Two graduate students in the computer science master's program at a large public research university labeled all comments. They achieved 85% consistency in their labeling, for the remaining 15% comments, they further discussed their discrepancies in annotation and reached agreement as well. We further feed the pre-trained model with the labeled data to fine-tune the parameters to better work for this task and context. For each comment category, we create a classifier, for a total of four classifiers. Lastly, we use the fine-tuned models

[10] https://github.com/google-research/bert

to predict all collected human comments. We ran the whole classification process on the Google Cloud Platform and used TPUs to perform all BERT-related processing.

Table 9. Variables and Descriptive Statistics

N	Variable	Description	Mean	S.D.	Min.	Max.
3,744	AutoMatedTasks$_{it}$	Continuous variables. It represents the number of moderation tasks delegated to AutoModerator on subreddit *i* in month *t*.	2.231	3.342	0	29
3,744	Num_User_Cmts$_{it}$	Continuous variable. It represents the number of comments that users created on subreddit *i* in month *t*.	89,351.66	3581 82.9	0	519, 2588
3,744	Num_Mod_Participation$_{it}$	Continuous variable. It represents the number of comments that human moderators created on subreddit *i* in month *t*.	322.119	668.175	0	6,542
3,744	Num_Mod_Casual$_{it}$	Continuous variable. It represents the number of casual comments that human moderators created on subreddit *i* in month *t*.	215.862	446.524	0	3,711
3,744	Num_Mod_Moderation$_{it}$	Continuous variable. It represents the number of moderation-related comments that human moderators created on subreddit *i* in month *t*.	106.257	318.407	0	4,737
3,744	Num_Mod_Policing$_{it}$	Continuous variable. It represents the number of policing type of comments that human moderators created on subreddit *i* in month *t*.	63.882	247.128	0	4,452
3,744	Num_Mod_Explanation$_{it}$	Continuous variable. It represents the number of explanation type of comments that human moderators created on subreddit *i* in month *t*.	80.166	282.480	0	4,460
3,744	Num_Mod_Suggestion$_{it}$	Continuous variable. It represents the number of suggestion type of comments that human moderators created on subreddit *i* in month *t*.	84.352	256.130	0	4,634
3,744	Num_Mods$_{it}$	Continuous variable. It represents the number of active human moderators on subreddit *i* in month *t*.	3.827	4.440	0	52

Table 8 shows the classification results of the four classifiers. We can see that, on average, we have achieved above 90% on the F1 score in four classification metrics. Notably, the policing classifier has reached 97.34% accuracy, 96.99% precision, 95.68%

recall, and 96.33% F1 score. Compared to typical NLP classification tasks, our BERT-based classifiers have achieved a satisfying prediction performance.

Lastly, with all collected data, we construct a series of variables for the corresponding data analyses. We aggregate our data at the subreddit-monthly level. We used the number of labeled comments created by human moderators to measure their various types of engagement. In addition to the data mentioned above, we also collect regular users' comments from each subreddit and use it as the proxy for user engagement in that community. Table 9 depicts detailed descriptions and summary statistics of all measures in this study.

3.4 Empirical Analyses and Results

3.4.1 Identification Strategy and Main Model

Considering the automation of moderation tasks is a shock to the community and that communities adopted bots for moderation at different times, we apply Difference-in-Differences (DiD) as our identification strategy. Specifically, the treatment group comprises communities that automated moderation tasks, whereas the control group includes communities that relied entirely on volunteer human moderators at each observational point. The same approach is applied in prior literature studying the impacts of ride-sharing entry (Greenwood and Wattal 2017; Burtch et al. 2018; Babar and Burtch 2020) and technology adoption (Tan and Netessine 2020). By comparing human moderators' activities of the treatment group and the control group, we are able to estimate how the integration of bot moderators to the platform governance affects volunteer human moderators' engagement.

Formally, our empirical model is presented in Model (1). $ModParticipations_{it}$ denotes the dependent variable, volunteer moderators' activities on the focal community. $After_{it}$ is the key independent variable. It is a binary variable, and the value of 1 means the moderation comment was generated after the subreddit implemented AutoModerator. Additionally, each human moderators' activity may be subject to the changes in the number of active human moderators and the number of submissions that their community received. Thus, we add two control variables—the number of active human moderators (i.e., Num_Mods$_{it}$) and the number of user comments (i.e., Num_User_Cmts$_{it}$) in respective subreddits—to the analysis. Another source of empirical concern is the unobserved intrinsic difference among communities and temporary shocks on the platforms. Therefore, we add the subreddit fixed effect θ_i and monthly fixed effect δ_t to control the unvarying subreddit and time impacts. ε_{it} is the error term.

$$ModParticipations_{it} = \alpha + \beta \times After_{it} + \gamma \times Controls_{it} + \delta_t + \theta_i + \varepsilon_{it} \quad (1)$$

3.4.2 Main Results

Table 10 shows the main estimation results. We found that the automation of moderation shows no effect on human moderators' overall participation in each subreddit. After we break down users' overall participation as moderation-related and informal interaction, the impact of moderation automation on these two aspects of moderator activities remains insignificant.

We further investigate the detailed changes in human moderators' activities. We perform our analysis again with three moderation focused measures (i.e., policing, explanation, and suggestions). Table 11 reveals several interesting findings. First, we can

see that the automation of platform governance encourages human moderators to perform more corrective behavior. Their policing activities increase by 20.2% after their community delegated moderation tasks to AutoModerator. Second, more supportive moderator behaviors also emerged. The explanation type of comments increases by 14.9%. The estimates of our control variables also align with real-world experience. When there are more user submissions and more active moderation team members, moderators' activities will also increase accordingly. These results reduce the concerns that our empirical analysis model may suffer from multicollinearity.

Table 10. The Impact of AutoModerator on Moderations by Human Moderators

Variables	$Num\_Mod\_Participation_{it}$	$Num\_Mod\_Casual_{it}$	$Num\_Mod\_Moderation_{it}$
After	-0.0566 (0.0676)	-0.0570 (0.0638)	0.0554 (0.0754)
Num_User_Cmts	0.183*** (0.0333)	0.175*** (0.0370)	0.137*** (0.0365)
Num_Mods	2.195*** (0.0751)	1.911*** (0.0746)	1.813*** (0.0683)
Constant	-0.679** (0.280)	-0.602* (0.325)	-0.953*** (0.311)
Subreddit fixed effect	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Number of subreddits	156	156	156
N	3,744	3,744	3,744
R-squared	0.670	0.604	0.555

Note: (1) Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; (2) For all dependent variables and the number of commenters, we use log transformation (i.e., log(x+1)) to accommodate skewed distribution and zeros.

Table 11. Human Moderation Breakdown

Variables	$Num\_Mod\_Policing_{it}$	$Num\_Mod\_Explanation_{it}$	$Num\_Mod\_Suggestion_{it}$
After	0.202**	0.149*	0.0745
	(0.0909)	(0.0838)	(0.0702)
Num_User_Cmts	0.110***	0.119***	0.108***
	(0.0364)	(0.0374)	(0.0353)
Num_Mods	1.442***	1.593***	1.727***
	(0.0890)	(0.0824)	(0.0724)
Constant	-1.061***	-0.985***	-0.781**
	(0.307)	(0.318)	(0.301)
Subreddit fixed effect	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Number of subreddits	156	156	156
N	3,744	3,744	3,744
R-squared	0.408	0.468	0.522

Note: (1) Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; (2) For all dependent variable and the number of commenters, we use log transformation (i.e., log(x+1)) to accommodate skewed distribution and zeros.

Overall, these results indicate that automating content moderation is positively associated with policing-related engagement in the community. Meanwhile, moderators also provide more support to their community members by offering more explanations about their community policies.

3.4.3 Incremental Impact of Moderation Automation

Our analysis in Section 4.2 considers moderation automation as one shock, and the shock intensity is the same for all communities. One of the unique aspects of our study is that the number of automated tasks could vary for different communities. To better differentiate the delegation extent and estimate the incremental impact of automation,

instead of using a binary indicator, we re-construct the measure of bot adoption with the number of automated tasks (Kummer et al. 2020). We next repeat the estimation of the impacts of moderation automation with this continuous measure. The results in Table 12 suggest that when communities delegate one more task to the bot moderator, human moderators' overall activities do not significantly change, but their moderation-related activities increase by 2.66%.

Table 12. Incremental Impact of Automated Moderation

Variables	$Num\_Mod\_Participation_{it}$	$Num\_Mod\_Casual_{it}$	$Num\_Mod\_Moderation_{it}$
AutomatedTasks	0.00137 (0.0105)	-0.000198 (0.0102)	0.0266** (0.0119)
Num_User_Cmts	0.179*** (0.0328)	0.171*** (0.0368)	0.134*** (0.0347)
Num_Mods	2.179*** (0.0753)	1.898*** (0.0746)	1.788*** (0.0650)
Constant	-0.653** (0.279)	-0.579* (0.325)	-0.920*** (0.297)
Subreddit fixed effect	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Number of subreddits	156	156	156
N	3,744	3,744	3,744
R-squared	0.669	0.604	0.558

Note: (1) Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; (2) For all dependent variable and the number of commenters, we use log transformation (i.e., log(x+1)) to accommodate skewed distribution and zeros.

Table 13 presents more detailed results for such moderation behaviors. Specifically, we find that one more automated moderation task will increase the policing, explanation, and suggestions by 5.15%, 4.6%, and 3.12%, respectively. The direction of these changes is consistent with our main analysis. Together, these results imply that bots

empower volunteers' role as community managers and enable them to achieve more community governance-related activities. Most importantly, while increasing their efforts at corrective behavior, human moderators also moved on to more advanced governance. They offer more support to their communities, including explaining their actions and giving more suggestions. These results complement the main analysis with more empirical evidence showing that the incremental regulation automation will lead to more voluntary, beneficial activities.

Table 13. Incremental Impact of Automated Moderation (Breakdown)

Variables	$Num\_Mod\_Policing_{it}$	$Num\_Mod\_Explanation_{it}$	$Num\_Mod\_Suggestion_{it}$
AutoMatedTasks	0.0515*** (0.0168)	0.0460*** (0.0149)	0.0312*** (0.0114)
Num_User_Cmts	0.112*** (0.0340)	0.118*** (0.0348)	0.105*** (0.0337)
Num_Mods	1.416*** (0.0851)	1.563*** (0.0781)	1.700*** (0.0685)
Constant	-1.036*** (0.288)	-0.950*** (0.297)	-0.747** (0.288)
Subreddit fixed effect	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Number of subreddits	156	156	156
N	3,744	3,744	3,744
R-squared	0.416	0.475	0.526

Note: (1) Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; (2) For all dependent variable and the number of commenters, we use log transformation (i.e., log(x+1)) to accommodate skewed distribution and zeros.

3.4.4 Additional Analyses and Robustness Check

We are aware that one of the main concerns about our study is the exogeneity of moderation automation. Specifically, there could be some hidden factors driving both

AutoModerator adoptions and human moderators' behavior change. We try to resolve this concern from both qualitative and quantitative perspectives. We interview with a selective number of senior moderators of our studied subreddits to learn about whether human moderators' engagement remain voluntary in both pre- and post-bot adoption periods. Moreover, we conduct several econometrical quantitative checks to test the validity of our analyses approach and the robustness of our results.

3.4.4.1 Interviews with Human Moderators

As we described earlier, we are focusing on the role of moderation automation on volunteer human moderators. It is possible that the human moderators make the decisions to implement bot moderators to facilitate pre-conceived community agenda. To understand whether human moderators' behavior changes due to pre-determined reasons, we interviewed 15 moderators who have volunteered on our studied subreddits for several years and learned whether their engagement is completely organic and voluntary. For example, a human moderator from a technology-related subreddit states that "There are literally no requirements asked of moderators." A moderator who has moderated more than 180 subreddits provided similar feedback by saying "None of the subreddits that I moderate have specific tasks assigned to specific moderators." When we asked our interviewee whether moderators' duties changed after their subreddit adopted AutoModerator, they responded "Nothing changed. AutoModerator give use more time to focus on reviewing the constant stream of content being submitted and overall managing the subreddit."

In sum, all moderators that we interviewed provide us consistent feedback. They all confirmed that moderators voluntarily contributed to the platform regulation during

both pre- and post- automation periods without particular agenda in mind. They also explained that AutoModerator helps volunteers handle easy and repetitive rule violations. Bots have better efficiency at maintaining their community environment and accommodate their subreddit growth. These interviews suggest that communities do not institutionally change moderators' roles after they adopted bot moderators.

3.4.4.2 Relative Time Model

Econometrically, to ensure our empirical approach is valid, we further examine the pre-treatment parallel trend assumption of our DiD model and test whether the difference between the treatment and control group remains the same over time before the treatment is assigned. We leverage the relative time model to examine the parallel trend assumption (Angrist and Pischke 2008; Greenwood and Wattle 2017; Burtch et al. 2018). Another benefit of using this approach is that the results will present a detailed picture of how bots affect human moderators' voluntary engagement over time. These results would help us understand the bot moderators' impact in both the short-term and long-term.

The empirical model for the relative time model is shown in Model (2). We create $RelativeMonth_{it}$ to represent the distance between the observational time point and the time when a community adopts the AutoModerator. We replace the variable $After_{it}$ with $RelativeMonth_{it}$ in Model (1) and perform our analysis again. Note that communities adopted AutoModerator at different times. To ensure the studied subreddits have enough observation in the pre- and post-treatment period, we choose subreddits that utilized bot moderation from 2013 July to 2014 June. Ninety-five subreddits remained in the analysis. We keep the observations in those subreddits from six months preceding the bot adoption

to six months after the adoption. Then, we construct the relative time model with the month preceding the bot adoption as the baseline.

$$ModParticipations_{it} = \alpha + \beta \times RelativeMonth_{it} + \gamma \times Controls_{it} + \delta_t + \theta_i + \varepsilon_{it} \quad (2)$$

We present the results in Figure 5. We first examine whether the parallel trend assumption holds before the moderation tasks are automated. Figure 5 shows that all observations in the pre-adoption period are not significantly different from the baseline under the 90% confidence level. This result supports the validity of applying the Difference-in-Differences approach as the parallel trend assumption is not violated. Most importantly, Figure 5 further presents the influences of automating moderation tasks on volunteer moderators' activities over time. We found a stable increase in moderators' policing behavior after delegating the moderation tasks to bots. With regards to two types of supportive behavior, explanations, and suggestions, the increases are more prominent in the first few months. Our earlier analyses exhibit that bots negatively affect moderators' casual engagement in the community. Interestingly, the evidence in Figure 5(d) depicts a clearer picture. It shows that automating moderation tasks encourage more informal interaction with users in the first three months. However, this effect decreases quickly afterward.

Overall, the relative time analysis validates our results and findings. It further shows us how automating moderation tasks affects human moderators' behavior dynamically. We find that bots lead to a positive increase in all aspects of volunteer moderators' activities. They accomplished more moderation and also had a more casual engagement with their community's users. However, supportive activities decrease compared to the previous activity level, and only the increase in policing behavior

persists in the long-term. This finding implies that automation effectively motivates volunteers to perform more basic corrective moderation, but that it has a limited influence on encouraging more supportive activities.

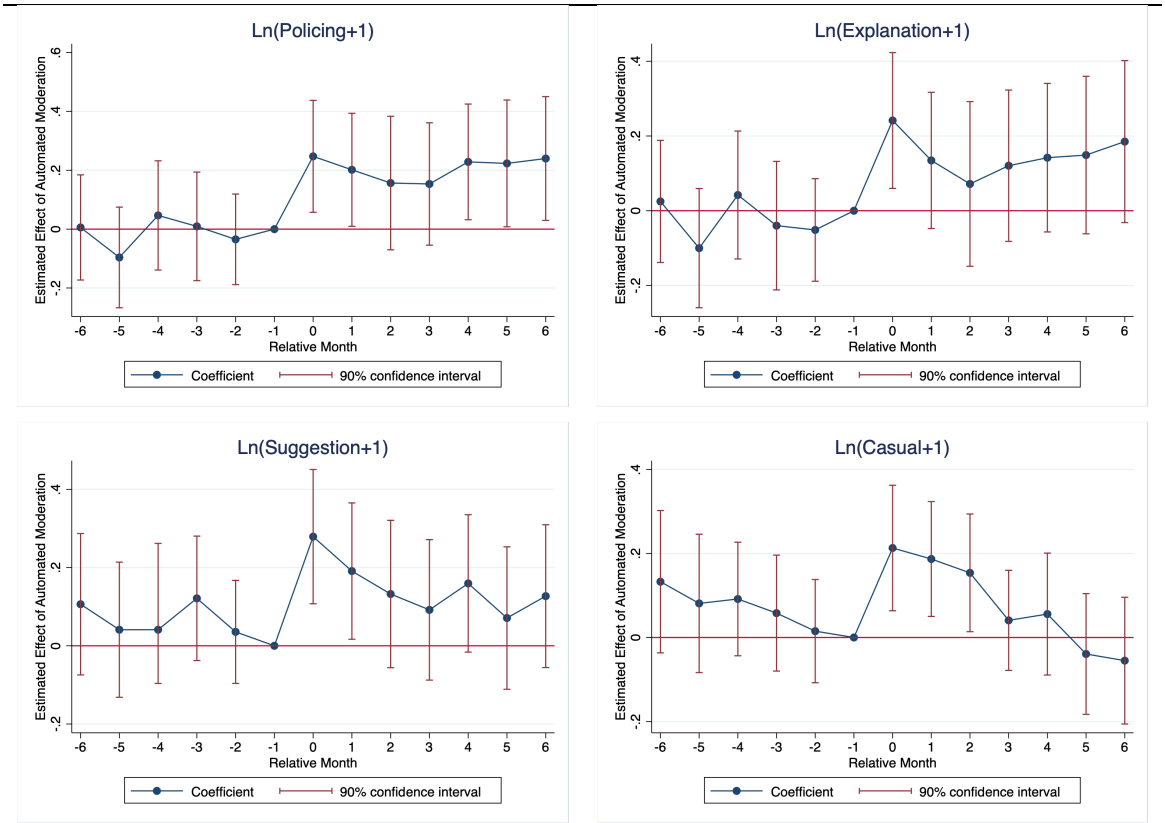


Figure 5. Estimation of Relative Time Model

3.4.4.3 Seemingly Unrelated Regression Model

Lastly, considering that human moderators' different types of participation might be correlated, we further apply the seemingly unrelated regression (SUR) model to account for the potential correlation among error terms of our regression models. Specifically, we take the subreddit and time fixed effects into account and re-estimate the impacts of task automation on four types of human moderators' activities simultaneously. Table 38 in the

Appendix reports the results. Overall, the estimations of the SUR model are consistent with our aforementioned findings.

3.5 Empirical Extension

3.5.1 Moderating Effect of Community Size

We perform two heterogeneity analyses to examine how community characteristics moderate the effects of bot moderators on human moderators' activities to further uncover some underlying mechanisms. The first community characteristic that we investigate is community size (Butler et al. 2014; Ren and Kraut 2014). Community size may play a role in the moderators' voluntary engagement for two reasons. First, community size implies community popularity and the resources (e.g., community content) available to their community members (Bulter et al. 2014). Moderators in popular communities receive more attention and higher prestige, as their moderation generates a more considerable impact. Such status and exposure will result in higher levels of commitment to community and engagement.

Second, compared to subreddits with fewer members, large communities usually have more urgent moderation needs to accommodate the massive amounts of submissions and diverse community requirements. Volunteer moderators in popular communities are expected to meet growing numbers of user requests and deal with a large corpus of user-generated content. When algorithm-based moderation tools become available, the augmentation impact will be more prominent for moderators from popular communities.

To test the moderating impacts of community size, we proxy the community size using the number of commenters in a subreddit as of the month when AutoModerator was adopted. Next, we split all studied subreddits into two groups and construct a binary variable $PopSubreddit_{it}$ to denote subreddits with relatively more members. We add this term and an interaction term between $PopSubreddit_{it}$ *and* $After_{it}$ into Model (1). Table 14 presents the estimation results. With the estimation of the automation decision (i.e., $After_{it}$) and the interaction term (i.e., $After \times PopSubredd_{it}$), we note that automating moderation tasks does not change moderators' policing and explanation activities in small communities but positively affect voluntary moderation in larger communities. Moreover, we also find that usage of bots oppositely affects moderators' suggestion activity in popular and non-popular subreddits. There was a 14.9% decrease in moderator suggestions on the small subreddits, and a 29.4% increase on the popular communities. With regard to the informal interaction with community members, we see a 16.2% decrease in both large and small communities. This empirical evidence suggests that machines weaken moderators' role as regular community members regardless of community popularity.

Overall, our results suggest that the human and moderation relationship is subject to the community size. On small communities, bots substitute volunteer moderators as evidenced by the decreased moderation-related and informal interaction. In contrast, bots exhibit positive impacts on moderators from large communities by augmenting their role as community managers. Such differences may be caused by the attractiveness of communities and the influence of moderators. When a community has a larger user base and more engaging content, moderators have greater impact and prestige on the platform.

They also face more extensive needs in moderating community content appropriately and promptly. Therefore, algorithm assistance enables these volunteers to stick with their moderator role and even achieve better performance, rather than leading them to disengage from the community.

Table 14. The Moderating Effect of Community Size

	Moderator Role			User Role
Variables	Num_Mod_Policing	Num_Mod_Explanation	Num_Mod_Suggestion	Num_Mod_Casual_Talk
After	-0.0690 (0.0930)	-0.122 (0.0941)	-0.149* (0.0889)	-0.162* (0.0882)
After × PopSubreddit	0.536*** (0.167)	0.537*** (0.152)	0.442*** (0.130)	0.207 (0.129)
Num_User_Cmts	0.114*** (0.0342)	0.123*** (0.0353)	0.112*** (0.0331)	0.177*** (0.0362)
Num_Mods	1.428*** (0.0902)	1.579*** (0.0836)	1.716*** (0.0733)	1.906*** (0.0744)
Constant	-1.073*** (0.289)	-0.997*** (0.300)	-0.791*** (0.280)	-0.606* (0.318)
Subreddit fixed effect	Yes	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
Number of subreddits	156	156	156	156
N	3,744	3,744	3,744	3,744
R-squared	0.419	0.478	0.529	0.606

Note: (1) Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; (2) For all dependent variable and the number of commenters, we use log transformation (i.e., log(x+1)) to accommodate skewed distribution and zeros.

3.5.2 Moderating Effect of Scope of Work

We next perform the second heterogeneity test on scope of work. Similar to rules and culture in organizations, community guidelines specify its identity and norms. These

guidelines provide quick guidance for users, especially newcomers, about the community focus and participation requirements. Meanwhile, they also specify volunteer moderators' scope of work and serve as a reference for them while governing their community.

We consider how scope of work will moderate the impacts of bots on human moderators' behavior for two reasons. First, when a community specifies a broad scope of work in governance, moderators will have less ambiguity in their role and better understand their responsibilities. Prior literature in the organizational context has examined that role clarity positively affects the employee's job commitment and satisfaction (Lyons 1971; Donnelly and Ivancevich 1975; Hassan 2013). We predict that when algorithms automate some routine moderation tasks, moderators from subreddits with more community guidelines are more likely to maintain their continued efforts and take care of other moderation tasks that cannot be automated by bot moderators. Second, clear and comprehensive scope of work officially document more specific circumstances that volunteers may run into in their moderation and the corresponding solutions. It provides volunteers more organizational support to perform moderations and justify their actions. Thus, it becomes easier for moderators to conduct advanced community governance, such as explaining their moderation decisions and recommending other available resources. Emotionally, detailed and thorough rules will also protect volunteers from having unnecessary conflicts with their members due to unclear and subjective moderation.

To test the moderating effect of a scope of work, we search all studied subreddits' historical pages during our observational period.[11] Then, we scrape the public community rules on these pages. We measure scope of work using the number of rules posted on a community. The more community rules, the broader a scope of work. We split the studied subreddits into two groups using the median value of the number of community rules and then label subreddits with more rules as group with a broad scope of work (i.e., *BroadSOW*). Next, we add the new term, *BroadSOW*, and an interaction term between *BroadSOW* and *After* into Model (1). See Table 15 for the results. The estimation is consistent with what we expected. Bots only show positive impacts on volunteers' moderation engagement in subreddits with more moderation tasks. In contrast, communities with fewer rules experience the same or even decreased volunteer effort in managing their communities after delegating tasks to bots.

Interestingly, from the two heterogeneity analysis results, we can see that community size and a scope of work present similar moderating effects on bot adoption. One of the natural concerns for these results is whether these two measures capture the same community characteristics. For example, larger communities are more likely to create more detailed community rules. If so, the moderating effects of a scope of work can be attributed to their community size. To test this concern, we perform another analysis by adding two interaction terms into the same model. We present our results in Table 16. The estimation on the two interaction terms suggests that both community size and a scope of work still independently moderate the impacts of regulation automation.

[11] We use https://archive.org/web/ in our study to find out historical subreddit pages.

Meanwhile, both factors show no effect on volunteers' informal interaction with their community members. As bot-assisted moderation becomes available and volunteers' moderation activities increase, their casual engagement drops.

Table 15. The Moderating Effect of Scope of Work

	Moderator Role			User Role
Variables	Num_Mod_Policing	Num_Mod_Explanation	Num_Mod_Suggestion	Num_Mod_Casual_Talk
After	-0.0735 (0.112)	-0.138 (0.104)	-0.174* (0.0910)	-0.128 (0.0790)
After × BroadSOW	0.557*** (0.170)	0.580*** (0.154)	0.504*** (0.128)	0.144 (0.127)
Num_User_Cmts	0.108*** (0.0327)	0.116*** (0.0337)	0.106*** (0.0322)	0.175*** (0.0366)
Num_Mods	1.406*** (0.0874)	1.556*** (0.0806)	1.694*** (0.0703)	1.902*** (0.0736)
Constant	-1.005*** (0.276)	-0.927*** (0.286)	-0.730*** (0.275)	-0.587* (0.322)
Subreddit fixed effect	Yes	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
Number of subreddits	156	156	156	156
N	3,744	3,744	3,744	3,744
R-squared	0.419	0.480	0.531	0.605

Note: (1) Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; (2) For all dependent variable and the number of commenters, we use log transformation (i.e., log(x+1)) to accommodate skewed distribution and zeros.

Overall, our results suggest that community needs for moderation are the driving factors for the increased voluntary engagements by human moderators after some moderation tasks are delegated to bots. These needs can come from the external factor (i.e., the community size). Communities themselves can also generate strong moderation

needs by enlarging and specifying moderators' scope of work. In the subsequent section, we will further discuss the managerial implications of the results.

Table 16. The Moderating Effect of Community Size and Scope of Work

	Moderator Role			User Role
Variables	Num_Mod_Policing	Num_Mod_Explanation	Num_Mod_Suggestion	Num_Mod_Casual_Talk
After	-0.223**	-0.285***	-0.293***	-0.194**
	(0.0896)	(0.0866)	(0.0874)	(0.0899)
After × PopSubreddit	0.412**	0.406**	0.326**	0.181
	(0.188)	(0.172)	(0.140)	(0.136)
After × BroadSOW	0.439**	0.463***	0.410***	0.0917
	(0.191)	(0.175)	(0.140)	(0.134)
Num_User_Cmts	0.112***	0.120***	0.109***	0.177***
	(0.0321)	(0.0332)	(0.0313)	(0.0362)
Num_Mods	1.403***	1.553***	1.692***	1.901***
	(0.0880)	(0.0813)	(0.0707)	(0.0736)
Constant	-1.026***	-0.948***	-0.747***	-0.597*
	(0.271)	(0.281)	(0.265)	(0.318)
Subreddit fixed effect	Yes	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
Number of subreddits	156	156	156	156
N	3,744	3,744	3,744	3,744
R-squared	0.425	0.485	0.535	0.606

Note: (1) Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; (2) For all dependent variable and the number of commenters, we use log transformation (i.e., log(x+1)) to accommodate skewed distribution and zeros.

3.6 General Discussion

3.6.1 Key Findings

In recent years, the need for and engagement in digital content has grown exponentially. Digital platforms have started widely adopting algorithms to offload human moderators

from ever increasing burden of routine moderation tasks. As platforms are moving to the technical mode of platform governance with help from algorithms, researchers have raised concerns of platform de-humanization and debate whether algorithms would lead volunteer human moderators to reduce their community contributions. In this study, we collect moderation records from Reddit and investigate the impact of machine-powered governance on volunteer human moderation. With data collected from 156 subreddits, we found that delegating moderation to machines augments volunteer moderators' role as community managers. Human moderators present more moderation-related engagement, including both corrective and supportive interactions with their community members. Notably, our results indicate that such effects manifest among communities with large user bases and detailed community guidelines, suggesting that community needs for moderation is the driving factor for volunteer moderators' increased contributions.

3.6.2 Theoretical Implications

Our work contributes to three streams of research. First, we contribute to the digital platform governance literature in the field of Information Systems (Van Alstyne et al. 2016). Prior work has investigated human-centered community governance and identified its positive impact on members' commitment and contribution (Ren and Kraut 2014). However, few studies have considered the role of algorithms in community governance. Our work investigates a novel form of platform regulation, which involves efforts from both humans and machine in content moderation. Moreover, our study further digs into human moderators' behavior and examines how implementing algorithm-based moderation tools affect these voluntary community managers' corrective, supportive, and

informal engagement. The computational method for engagement classification based on state-of-the-art machine learning will also offer a new approach to future studies in platform governance.

Second, our work also sheds light on the human-machine frontier. Human-machine collaboration has been widely studied from the traditional business sector (Dixon et al. 2020) to the online context (Luo et al. 2019). In particular, emerging studies in recent years have focused on the impacts of artificial intelligence on online contexts such as e-commerce (Schanke et al. 2021). Our study contributes to this research stream by adding empirical evidence on an increasingly important yet understudied area—content moderation. Our research particularly explores the relationship between machines and uncompensated human labor (i.e., community moderators). The results suggest that algorithm-based moderation augments the volunteer human moderators' role as community managers. Delegating routine and tedious moderating work to algorithms will lead volunteer human moderators to perform more corrective and supportive engagement in their community.

Third, our work also contributes to content moderation in the literature on computer-supported cooperative work. There has been increasing attention surrounding the sustainability and impact of human-machine collaboration in content moderation in recent years (Zheng et al. 2019; Dosono and Semaan 2019; Matias 2019 b). More specifically, there is a growing call for re-humanizing platforms in the algorithmic era of platform regulation. To the best of our knowledge, our work is among the first that provides quantitative, empirical evidence showing that machines enable volunteer humans to make more moderation-related contribution. Moreover, our empirical

extensions also show that such impacts are subject to community needs for moderation. Specifically, subreddits with larger user bases and subreddits with detailed guidelines experience an increase in human moderators' voluntary engagement.

3.6.3 Practical Implications

Our results first dispel concerns from community managers that applying machines to platform regulation will drive volunteer moderators away from their community and cause less community engagement. In contrast, equipping volunteer moderators with more technological moderation tools will augment their role as community managers. Machines empower these volunteer moderators to perform more comprehensive community governance, which may require subjective interpretation and a better understanding of the context. Most importantly, while achieving more comprehensive corrective engagement, volunteer moderators contribute more supportive moderation to their community. This result resolves concerns of the de-humanizing effects of algorithmic regulation as machines will stimulate more positive engagement from volunteer moderators. Our finding also imply that compared to governance models that solely rely on human moderation, human-machine collaboration brings more sustainability to the volunteer moderation team and in turn helps the community grow.

Our study suggests that not all communities would benefit from moderation automation. Popular communities with a massive user base will enjoy the increase in volunteer moderators' contributions. However, for small communities with limited influence and attractiveness, platform practitioners can encourage moderators to establish clear and detailed community guidelines. By enlarging and specifying moderators' scope

of work with more detailed community guideline, automation can also stimulate more voluntary contributions in the community.

Our work further suggests that platform practitioners need to be careful about the decreasing informal conversations between moderators and community members when regulation is automated. Interpersonal communication is a critical factor in enhancing bond-based attachment (Ren et al. 2012). Ideal platform governance also needs to balance formal regulation and informal social control (Williams 2007). Our results indicate that communities with relatively small user bases and non-entertainment related communities have shown a more significant decrease. Considering the importance of interpersonal interaction to communities, platform practitioners should evaluate the impact of reduced informal engagement and guide volunteer moderators to avoid potential adverse outcomes.

3.6.4 Limitations and Future Work

We understand that there are some drawbacks to our research. First, as documented in prior literature (Jhaver et al. 2019 c), moderation records on Reddit are stored in three ways: public moderation comments, submission flair, and private conversations with content creators. Given the data availability, our data only comes from public moderation records. Therefore, we may miss some moderations recorded in submission flair and private messages. However, we do not consider this to be a severe issue in our study for several reasons. First, prior literature (Jhaver et al. 2019 c) found that more than 80% of moderation records are present in the form of public comments. Related work also utilized this data collection approach. Second, public moderation records represent an

essential dimension of platform governance—transparency. Transparent moderation is an effective way to resolve moderating conflicts and educate users about community norms. As calls for transparent content moderation continue to be raised (Jhaver et al. 2019 c), public moderation will become the dominant regulation approach on Reddit. Therefore, studying the impacts of machines on moderators' transparent moderation is also meaningful for platforms. Meanwhile, we also open this avenue for future researchers to extend the work if the comprehensive moderation records are accessible. It would be interesting to investigate the influences of machines on moderators' choice of different communication channels.

Second, our study only investigates the human-machine collaboration from the angles of changes in volunteer human moderators' community engagement. We did not consider the new effort and time that human moderation spent creating and maintaining this collaboration (i.e., human-in-the-loop). For example, to assimilate bots into community governance, the moderation teams on Reddit need to have some tech-savvy members to design bots, fix algorithmic errors, and ensure the bots are aligned with updated community guidelines (Jhaver et al. 2019 a). However, these costs occur in the background, and are difficult for researchers to observe. In this study, we only focus on the changes in volunteers' moderation, the direct interaction with community members. Nevertheless, human-in-the-loop would be an intriguing topic for future research. It will generate valuable insights into moderation teams' work design and management.

Last but not least, while Reddit has become a giant online platform with a large user base, other platforms may have different platform regulation structures, and they may apply different approaches in integrating algorithms into daily content moderation

(Yu et al. 2020). Therefore, it is necessary to go beyond our research context and examine the external validity on other platforms. Achieving this goal requires understanding business logic for different platforms and effective approaches to access data. Therefore, we will open this avenue to future researchers and look forward to more insights into good human-machine regulation on other platforms.

We believe our study makes an initial step in understanding the impact of machines (particularly algorithms) on humans' voluntary content moderation, which plays an increasingly important role in our current and future digital life. We hope our study can bring more attention to this understudied area, and more future work could build on this study.

CHAPTER 4

DOES IDENTITY DECLARATION AMPLIFY OR ATTENUATE POLARIZATION IN ONLINE POLITICAL DISCOURSE?

4.1 Introduction

Political identity has become a critical social identity in the era of digital platforms (Greene 2004; Fowler and Kam 2007). It plays an important role in critical political decisions such as civic engagement, preferences about social policy, and vote choice (Bartels, 2002; Dimock et al. 2014). Meanwhile, it also commonly exists in daily political discussions and is situated in the center of the rising concern of political polarization. With the rapid development of technology, online discussion forum has become the dominant venue for people to engage in political discussion. Users constantly interact with people holding various perspectives with varying information about the other party. While online platforms provide flexible designs to enable users to disclose their political identity, it is critical for platforms to know how such disclosure affects political discourse.

Existing literature has studied the importance of identity and disclosure from various aspects. First, social identity and political science literature have revealed that identity would causally affect individuals' behavior (Gerber et al. 2008; Fowler and Kam 2007; Turner et al. 1987; Tuner et al. 1994). However, these work are conducted in the offline setting and lack empirical evidence in the online context. Second, researchers have examined the value of information disclosure in a variety of online contexts (Forman et al. 2008; Burtch et al. 2016; Pu et al. 2020). However, few work has been

paid to political identity and its impacts on online discourses. Lastly, rising attention has been paid to political polarization (Bail et al. 2018; Han and Hu 2021) in recent years. But few papers have studied polarization from the angle of identity salience in the online context. Therefore, more work is yet to be done to bridge these streams of research.

Political identity disclosure may have mixed impacts on user participation in online political discourse. On the one hand, political identity summarizes individuals' existing views and enables users to quickly identify others with similar political perspectives. Shared political identity can enhance the community attractions and strengthen community bonds (Ren et al. 2007) hence stimulating user participation. On the other hand, declaring political identity can also highlight different political views. For users with views different from the majority, the disclosure will highlight their status as the marginalized group and discourage them from participating in the subsequent discussions.

Moreover, the impacts of political identity disclosure can go beyond user participation and further affect the polarization in subsequent discussions. Declaring political identity increase the salience of in-group and out-group in online discourse. When explicitly declaring one's political stance, the minority views are more likely to stand out and receive more attention in a political discussion. The increased exposure to the different views will facilitate the view exchange and further less polarized conversation (Levy 2021; Greenstein et al. 2016). However, prior literature also indicates that disclosing political identity can backfire and amplify polarization. Due to the increased salience of political identity, users may adopt attitudes or behaviors aligned with their group norms (Gerber et al. 2010; Forman et al. 2008; Burtch et al. 2016).

Meanwhile, the strengthened distinction between in-group and out-group can also lead to more polarized opinions towards people with different views (Bail et al. 2018).

As such, we will seek to understand the implications of political identity disclosure for the extent and intensity of individual participation in online political forums. As a secondary question, we will examine the disclosure causally impacts the polarity of online political discourse. One of the most prominent empirical challenges of identity study is to disentangle the homophily from the causal impacts of identity declaration on behavior changes (Gerber et al. 2010). Our study takes advantage of a policy change on an online discussion community as an exogenous shock to study how the mandatory political stance disclosure impacts subsequent political discourse in the community. Specifically, in our study, we collect the data from Reddit, a large news aggregation and discussion website in the world. In August 2018, a center-right community named r/tuesday implemented a new flair policy requiring all participants to declare their political stance in their future content. Otherwise, users cannot participate in any discussions. While the shock happened to r/tuesday, other political discussion communities did not experience such policy change. Thus, these communities form a control group naturally.

Econometrically, we apply the Difference-in-Differences approach to estimate the impacts of the mandatory flair policy on user behavior in terms of participation and discussion polarization. Our results indicate that after disclosure, there is a significant drop in participation of new users in a post. Additionally, we observe increased polarization during the interaction between opposing views. Interestingly, users holding different political perspectives are affected disproportionately due to the mandatory

disclosure policy. The most significant decrease in engagement has been found on users with undeclared political stances. In contrast, the left-leaning users, the minority group, maintained a comparable amount of activity as before but used more slang and partisanship terms in their subsequent discourse.

Our research extends past research in at least three important ways. First, prior work has examined the impacts of identity disclosure in various online contexts (Forman et al. 2008; Lu et al. 2019; Pu et al. 2020). Our work contributes to this research stream by extending the research to political identity, an increasingly salient and common social identity in online platforms nowadays. Our results show that political identity disclosure discourages users, especially new users and users with an unclear political stance, from participating in online political conversations. Second, our work also contributes to the growing research stream on online polarization by examining the impacts of political identity on the polarity of interaction between people with opposing political views (Bail et al. 2018; Greenstein et al. 2016). Our results indicate that political identity can cause the overall political discourse to be more polarized. Third, our work contributes to the growing online content moderation literature (Jaidka et al. 2019; Matias 2019 b) by empirically examining the impacts of mandatory identity disclosure policy on overall community participant combination and environment. The results show that the policy will form a more homogeneous group but meanwhile cause more rule violations.

4.2 Literature Review

4.2.1 Identity Disclosure

Internet is created with an anonymous nature. In recent decades, more and more platforms start exploring the value of disclosing identity-descriptive information. Prior literature has widely documented the impact of identity disclosure in a variety of online contexts, such as e-commerce (Forman et al. 2008), crowdfunding (Burtch et al. 2016), social media (Cavusoglu et al. 2016; Kilner et al. 2005), auction (Lu et al. 2019), and Q&A forum (Pu et al. 2020). For example, Forman et al. (2008) found that on e-commerce platforms, public descriptive information would build a social norm and influence other users' disclosure decisions. Reviewers' disclosure of descriptive information can positively influence the perceived helpfulness of review and ultimately lead to increased product sales. Similarly, with the focus on crowdfunding platforms, Burtch et al. (2016) found that revealing users' campaigns positively influences the subsequent likelihood of visitor conversion and average contribution. Additionally, with data collected from user-generated content platforms, Pu et al. (2020) find that disclosing participant identity will inhibit user content generation on the focal community but increase their contribution to neighbor communities.

Despite these efforts, extant literature has paid little attention to political identity disclosure. Prior results cannot simply apply to the context of online political discourse, given the substantial differences between political identity and those studied in extant work. Prior literature mostly focuses on the disclosure of individual associated information (e.g., real name, gender, and location) in contrast to the anonymous online

environment. This type of identity mainly carries neutral information and can be used to uniquely identify content creators. In contrast, political identity is a type of social identity. It is associated with a particular group that shares the same political views (Greene 2004). Instead of delivering a message about who I am, political identity disclosure expresses content creators' general beliefs in political issues, and such beliefs may significantly affect their interactions with participants. Therefore, it is necessary to collect more empirical evidence to examine the role of political identity in online political discourses.

4.2.2 Online Platform Polarization

In recent years, a growing number of studies have investigated the interaction between people with different political ideologies, and these work show mixed findings. Bail et al. (2018) implemented an experimental intervention, repeatedly posting content in opposition to the perspective of subjects on Twitter users. Those authors observed that persistent exposure to opposing views had the unfortunate effect of inducing even greater polarization in opinions and beliefs. However, in other contexts, exposure to different sides of the voice can facilitate different political perspectives reach to the consensus and even lead to less polarized opinion towards the other groups. Greenstein et al. (2016) study the contribution to politics-related articles on Wikipedia, and they find that contributors tend to edit articles with slang opposing their views. Thus, the interaction among different views ultimately results in less segregated conversations and fewer biases on the whole platform. Another recent empirical evidence from social media also shows similar findings. Levy (2021) conducted a large-scale field experiment on

Facebook. Levy found that exposure to counter-attitudinal news can effectively decrease the negative attitude towards the opposite political party.

Overall, existing studies primarily study the causes of polarization from the angle of exposure to different perspectives. Although political polarization happens around people with different political beliefs, little work has investigated the polarization issue from the angle of political identity salience. Moreover, the contradicting results in prior literature also suggest contextual factors may significantly influence the polarity in the political conversation (Urman 2020). With data collected from the organic and widely used online political discourse venue, discussion forum, our work aims to fill this research gap by conducting empirical analyses.

4.3 Hypothesis Development

4.3.1 Political Ideology Disclosure on User Participation

Given that political identity reveals individuals' attitudes and general political beliefs rather than neutral personal information, it is critical to consider how the identity disclosure would affect the relationship between community members. The revelation of political identity is a process of social categorization. The descriptive-identity information enables users to identify like-minded users and users with opposing views. It further leads to the increased distinction between in-group and out-group. As a result, the relationship between groups may replace the inter-members relationship and become the dominant one.

The majority and minority groups may react differently to such change. First, the shared identity and increased distinction between in-group and out-group can enhance

their perceived common identity in the community (Ren et al. 2007) and strengthen their community attachment. Moreover, users who share the perspective with the majority are more likely to receive affirmation and positive feedback from the community. Due to the increased community attachment and affirmation, users with the majority views are more likely to engage more after their political stances are made known.

In contrast, disclosing political identity may pose the minority in a disadvantaged situation. Fewer users share the same political perspectives as them. Because their minority identity differentiates them from most users in the community and challenges the dominant opinions, they are more likely to stand out in online discourse and receive more negative feedback and counterargument in future discussions. Thus, declaring their political identity may discourage their future participation.

In addition to the changes in the inter-members relationship, identity disclosure can also add extra participation costs for users. Self-identifying requires users to reflect on their political views and affirm their identity choice before joining the discussion. Compared to users who are already deeply committed to their political ideology, the participation cost is higher for inexperienced users such as newcomers who need to compliance the community norms. The cost is also higher for users who are unsure about political identity or are unwilling to choose an identity. The enforced identity disclosure may make them experience more difficulty and feel vulnerable in future participation. It would further result in their decreased engagement in the community.

Given the potential changes in the inter-members relationship after disclosing users' political identity, we consider a wider range of community users would experience the negative influences of such disclosure. Therefore, we propose our first hypothesis:

H1: Disclosure of political stance will lead to decreased user engagement in the focal community

4.3.2 Political Stance Disclosure on Polarization

The participation change can further affect the polarity in the subsequent online political discourses. Continue our discussion about user participation; political identity disclosure can lead to a more homogenous environment. As the salience of shared identity and interaction with like-minded users increases, users are less likely to have conflicts with others within the community.

Moreover, given that disclosed identity will increase the distinction between in-group and out-groups, it can further increase exposure between different political views. Numerous social identity and political science literature have studied the interaction of conflicting groups and present mixed results. On the one hand, more exposure to the opposing group can break the echo chamber, facilitate the idea exchanges in a conversation and further form a better understanding of the other party. For example, by showing users news from the opposite political stance on Facebook, Levy (2021) found that such exposure will effectively decrease users' negative attitudes towards people leaning toward other political parties. As a result, their opinion becomes less extreme. Similarly, Greenstein et al. (2016) also show similar empirical evidence with data from Wikipedia. Their results suggest that the contributors have a stronger intention to edit the extreme slang from the opposing political side. Different voices will ultimately reach a consensus, and therefore the finalized article will be more neutral. These empirical

evidence shows that more idea exchange can manifest when users quickly identify the opposing opinions through the disclosed political stance.

On the other hand, cross-cutting exposure can also backfire and amplify the polarity in a political discussion. First, the increased interaction with the opposing view will challenge users' existing political beliefs and cause more cognitive dissonance (Bail et al. 2018; Pettigrew and Tropp 2006). When facing political divergence and counterargument, people tend to resort to the perceived differences between groups and strengthen their political beliefs, leading to a more unpleasant and polarized conversation (Nyhan and Reifler 2010; Taber and Lodge 2006). To resolve the conflict and reach the consensus requires extraordinary effort and patience, which are commonly absent in the online environment. Bail et al. (2018)'s field experiment provides empirical evidence showing that expose users to the opposing views sent by bots will result in more polarization.

Additionally, asking users to reveal their political identity can also influence users' internal thinking process and ultimately affect their future attitude and behavior (Gerber et al. 2010; De and Rizzi 2016; Oyserman and Dawson 2020). First, disclosing political stance is a reflection process that requires users to think over their political views and measure the alignment with their chosen identity. Once users commit to their ideological stance, the disclosure will enhance the salience of their political identity. It will further lead them to adopt more group-like attitudes and behavior and even develop into in-group favoritism and out-group derogation (Brewer and Brown 1998; Rogowski and Sutherland 2016). Gerber et al. (2010)'s field experiment in the offline setting

elegantly demonstrates that the strengthening partisan identity causally results in the shift in individuals' candidate preference and their evaluation of a salient political figure.

Taking the above aspects together, we believe that the impacts of political identity disclosure are more likely to be driven by the increased polarization. Thus, we propose our second hypothesis:

H2: Disclosure of political stance will lead to more polarized political discourse

4.4 Empirical Setting

4.4.1 Research Context

We collect our data from Reddit, a social news aggregation and discussion online platform. Reddit ranks as the seventh most-visited website in the US as of February 2021. It comprises over 130,000 active user-created communities called "subreddits" covering various topics such as news, politics, and video gaming. Political discussion has become a dominant activity on Reddit. Politics-related subreddits bring tremendous traffic to the platform as well as many controversial online discussions.[12,13]

Reddit is an ideal setting for our study, given its flexible platform design and highly autonomous community management. Like most social media, to participate in Reddit, users only need to register an account without revealing their true identity. However, Reddit enables users to display more personal information by adding textual and image flair. When they create content on Reddit, the platform will display user flairs along with their content. Many subreddits utilize this feature design to organize user

[12] https://www.forbes.com/sites/nicholasreimann/2021/01/08/reddit-bans-rdonaldtrump-subreddit/?sh=69bde1eb38b3
[13] https://www.npr.org/2020/06/29/884819923/reddit-bans-the_donald-forum-of-nearly-800-000-trump-fans-over-abusive-posts

participation. For example, a debate community, r/changemyview, uses flair to show user recognition in their community, and a science discussion subreddit, r/science, utilizes flair to signal a user's educational background and expertise.

Most importantly, on Reddit, all subreddits are created and managed by users. In addition to obeying the platform-wide guideline, each community can apply their unique subreddit rules to manage their user activity and meet their community needs better. This autonomous community structure enables us to study the variety and impact of platform policy. In this study, we take advantage of a policy change in a community to study the effect of identity declaration on participation and polarization in online discourse.

Specifically, our study focuses on r/tuesday, a center-right community, and its mandatory flair disclosure policy. [14] On July 31st, 2018, r/tuesday, announced a new policy requiring all users to add user flair disclosing their political stance in their future content. If users fail to add the flair, they cannot create any content in the subreddit. This community-wide policy change is exogenous to r/tuesday users. Meanwhile, users from other politics-focused communities did not experience such policy change, and therefore these communities form a control group naturally. Moreover, this flair policy also provides us with a great opportunity to accurately observe users' political leaning, a commonly hidden but increasingly important societal background of a user. Prior literature mainly infers online users' political ideology or partisanship based on their social network (Demszky et al. 2019), which may inherently contain large measurement errors. In our study, we are able to observe users' political stances based on their self-disclosed identities. Such information gives us better estimations in our research and

[14] https://www.reddit.com/r/tuesday/

enables us to disentangle the underlying mechanism by exploring the heterogeneity in users with different political opinions.

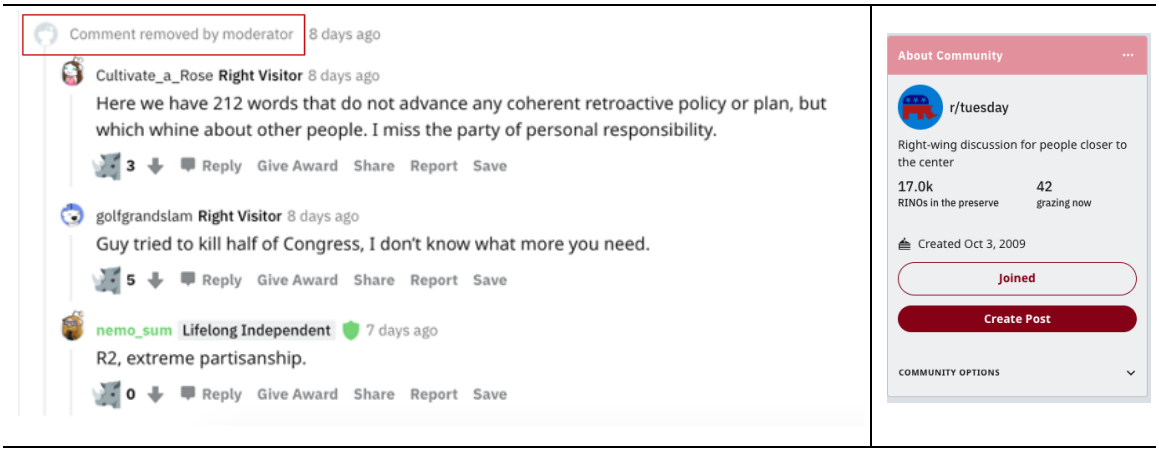


Figure 6. Screenshots of r/tuesday on Reddit and a Removal Message

4.4.2 Data Collection and Measures

In this research, we take advantage of exogenous community policy change on r/tuesday and use it as our focal study group. Meanwhile, we choose r/moderatepolitics as the control group because this subreddit is created around the same time as r/tuesday, and they both focus on moderate political discourse. We collect all content generated on r/tuesday and r/moderatepolitics from February 1st, 2018, to January 31st, 2019, using Reddit API. This observation window covers all community activity from six months preceding the flair policy implementation to six months after that.

We focus on user commenting behavior because commenting is the primary way to participate in political discourse on Reddit, and all comments are made under a post. Therefore, we organize our data at the post level and measure user participation by (1) the number of comments (i.e., Num_cmts_i) and (2) the number of unique commenters (i.e., Num_cmters_i) in a conversation. Additionally, we specifically look at the participation of

new users, given that new users are the driving force for new perspective creation (Burtch et al. 2020). Prior literature (Matias 2019 b) also indicate that newcomers behave differently and are more sensitive to the policy changes in the community. Thus, we further add two measures, (3) *Num_cmts_by_new_users*$_i$ and (4) *Num_new_cmters*$_i$, to additionally measure new users' engagement.

Table 17. Variables and Descriptive Statistics

N	Variable	Description	Mean	S.D.	Min.	Max.
6,327	Num_cmts$_i$	The number of comments under post *i*.	18.091	47.042	0	1223
6,327	Num_cmters$_i$	The number of users who participated in the discussion under post *i*.	6.972	10.321	0	295
6,327	Num_cmts_by_new_users$_i$	The number of comments by new users under post *i*.	1.862	11.310	0	811
6,327	Num_new_cmters$_i$	The number of new users who participated in the discussion under post *i*.	0.895	3.143	0	205
6,327	Slang_cmts$_i$	The number of comments contains political slang under post *i*.	0.310	1.797	0	98
6,327	Partisanship_cmts$_i$	The number of comments contains partisan words under post *i*.	2.683	7.920	0	253
6,327	Poster_stance$_i$	Categorical variable, 0 if post *i* is created by a left-leaning user; 1, if post *i* is created by a right-leaning user; 2, if post *i* is created by a user with unclear flair.	1.182	0.782	0	2
6,327	Discussion$_i$	Dummy variable, 1 if post *i* is a Q&A discussion; 0, if post *i* links to news or articles.	0.086	0.280	0	1
6,327	Num_Posts_Week$_{it}$	The number of posts created in week *t* on subreddit *i*.	75.504	30.908	15	143

Regarding polarization, extant literature has found the polarization is closely associated with language usage between the opposing groups (Gentzkow and Shapiro

2010; Greenstein et al. 2016; Gentzkov et al. 2019) under different environments (An et al. 2019). Thus, we count the slang and partisanship term usage in the collected comments. Specifically, we organize a list of political slangs and partisanship words pointing to other groups with different political stances. Then, we count the number of comments containing the slang or partisanship terms under a discussion thread and use these results as additional measures for discourse polarization. Text-based measures enable us to validate the polarization results and offer us another angle to investigate the changes in visible conversation. Table 17 displays all variables and descriptive statistics. We will present our empirical model and results in the next section.

4.5 Empirical Analyses

4.5.1 Model

We use Model (1) to formally examine the impacts of identity declaration on the subsequent political discourse. In this model, the variables of interest are user participation and discussion polarization. Our key independent variable is the interaction term between *Treated* and *After*. A series of control variables are included in our study. First, considering the effort to join a conversation, we control the post type (*Discussion*) by differentiating a post as article-triggered discussion or discussion-triggered discussion. Moreover, we distinguish conversations by the creator's political stance (*Post_by_Left* and *Post_by_Right*) to further control starting opinion of a conversation. We use posts created by users with an unclear political stance as the baseline in our empirical analyses. Next, to control the influence of the available discourse in a given week, we further include the number of posts in that week (*Num_Posts_Week*) into our model so that we

can eliminate the changes driven by the conversation availability. We further include the number of comments (*Num_Cmts*) received by a post for the discourse polarization analysis. Finally, we add month dummy variables into our model to control the unobservable temporary and seasonal impacts.

$$Participation_i/Polarization_i = \alpha + \beta_1 \times After_i + \beta_2 \times Treated_i \times After_i + Control_{it} + \varepsilon_i \quad (1)$$

4.5.2 Main Results

We apply the fixed effect model to estimate the impact of identity declaration on subsequent user engagement. The user engagement results are displayed in Column (1) – Column (4) of Table 18. According to the estimation in the interaction term, we find no significant changes in the overall user engagement after the community implements the identity disclosure policy. However, the number of new users and new users' comments decrease by 19.5% and 24.4%, respectively, in the subsequent period. These results suggest that identity declaration would discourage new users from participating in the online discourse.

We use negative binomial regression to estimate the results for discourse polarization because all dependent variables are count numbers, and the variance of these variables is greater than their mean. We restrict our analysis to conversations that received at least one comment. Column (6) shows that there is a slight decrease in the partisanship terms usage. Additionally, the estimations of other control variables are quite as expected. For example, the results show that when more discourses are available in a week, the user engagement and the polarization in a post reduce.

Table 18. The Main Effect on User Participation and Polarization

	Participation				Polarization	
Variables	(1) Num_cmts	(2) Num_cmters	(3) Num_cmts_by_new_users	(4) Num_new_cmters	(5) Slang_cmts	(6) Partisanship_cmts
Treated	-	-	-	-	0.570***	0.400***
					(0.175)	(0.0838)
After	0.822	0.685**	0.318	0.258	0.132	0.00109
	(0.141)	(0.0455)	(0.0737)	(0.0485)	(0.171)	(0.0726)
Treated x After	-0.542	-0.456	-0.244*	-0.195**	-0.123	-0.138*
	(0.114)	(0.0865)	(0.0221)	(0.0106)	(0.170)	(0.0704)
Post_by_left	0.0160	0.0328	-0.0270**	-0.0105	0.0201	0.129***
	(0.107)	(0.0514)	(0.000970)	(0.0134)	(0.101)	(0.0402)
Post_by_right	-0.109	-0.0729	-0.0686*	-0.0504**	0.0109	0.0719*
	(0.145)	(0.0642)	(0.00789)	(0.00275)	(0.0935)	(0.0389)
Discussion	1.220	0.867	0.628	0.448	0.117	0.204***
	(0.679)	(0.487)	(0.141)	(0.0933)	(0.0985)	(0.0391)
Num_Posts_Week	-0.543*	-0.416**	-0.167***	-0.134***	-0.0664	0.0965*
	(0.0499)	(0.00718)	(0.00180)	(0.00144)	(0.123)	(0.0499)
Num_Cmts	-	-	-	-	0.922***	0.921***
					(0.0320)	(0.0124)
Constant	3.963**	3.100**	1.178**	0.929**	-4.503***	-3.380***
	(0.219)	(0.101)	(0.0302)	(0.0383)	(0.563)	(0.230)
N	6,327	6,327	6,327	6,327	5,109	5,109
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.117	0.117	0.083	0.085	-	-

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Overall, our main results indicate that political identity disclosure decreases user engagement. In the following section, we will further disentangle the impacts of identity declaration by exploring the moderating effect of the discourse type and participants' political leaning.

4.6 Mechanism Exploration

One of the biggest advantages of our study is that users' political stances in the studied subreddit are observable. This data availability enables us to explore the underlying mechanism from both conversation type and participant aspect without including the potential measurement error from user identity inference. Given that user political ideology is only fully observable in the focal subreddit, in this section, we conduct a series of analyses with the data from the focal subreddit. Our analyses further reveal how identity declaration impact different types of online discourse and how such policy influences users at various locations in the political spectrum.

4.6.1 How Do User Behavior Change on Political Article versus Discussion?

Based on the cause that stimulates the conversation, we differentiate an online political discourse as article- or discussion-triggered content. Participant characteristics and focus of these two types of political discourse might be different. News- or article-triggered discourse requires users to digest the content first before they participate in the conversation. Therefore, users with a stronger political interest and more mature political views are more likely to consume such content and contribute to the discourse. In contrast, discussion-triggered discourse does not require users to spend extra effort reading additional content. Users can make comments based on their own opinions. Thus, it is easier to engage a wide range of users, and the discourse is more likely to get intense. When the identity declaration becomes mandatory, the user engagement and polarization in the article- and discussion-triggered content can be disproportionally affected. From the information process perspective, in the discussion environment that normally contains

massive and diverse opinions, people tend to rely on peripheral route processing and make their judgments based on simple source cues such as participants' political identity (Forman et al. 2008; Petty et al. 1998). Therefore, we first explore the underlying mechanism by examining the heterogeneity in the discourse type.

Table 19 shows the results. The estimation of the interaction term suggests that the overall engagement in discussion-focused discourse is not significantly different from the engagement in article-focused ones. However, we observe the heterogeneity impacts in the new user engagement. From Columns (3) and (4) in Table 19, we can see that new users' comments in discussion-triggered posts are 21.4% higher than those in political articles. Combined with the main effect, identity declaration decreases new users' participation in the article-related posts but increases their engagement in discussion-related content.

Regarding the polarization effect, we did not observe the significant difference between article- and discussion-triggered discourses. Overall, Table 19 implies that the mandatory ideology disclosure shifts new users' engagement from political articles to discussions. Therefore, it would be a good strategy for community moderators to have discussion-type posts to engage more newcomers.

Table 19. The Moderating Effect of Online Discourse Type

	Participation				Polarization	
Variables	(1) Num_cmts	(2) Num_cmters	(3) Num_cmts_by_new_users	(4) Num_new_cmters	(5) Slang_cmts	(6) Partisanship_cmts
After	-0.0907 (0.148)	-0.0178 (0.111)	-0.173** (0.0834)	-0.132** (0.0615)	0.282 (0.416)	0.146 (0.167)
Discussion	1.974*** (0.121)	1.385*** (0.0817)	0.641*** (0.0860)	0.430*** (0.0600)	0.00238 (0.230)	-0.0227 (0.105)
After x discussion	-0.203 (0.184)	-0.0848 (0.120)	0.214* (0.123)	0.185** (0.0848)	0.316 (0.292)	0.182 (0.121)
Post_by_left	-0.0318 (0.0835)	0.0521 (0.0612)	-0.000787 (0.0442)	-0.00254 (0.0326)	0.212 (0.214)	0.138 (0.0892)
Post_by_right	-0.200** (0.0823)	-0.0723 (0.0605)	-0.0502 (0.0433)	-0.0380 (0.0320)	0.0608 (0.216)	0.0333 (0.0877)
Num_Posts_Week	-0.310*** (0.107)	-0.285*** (0.0803)	-0.0853 (0.0559)	-0.0718* (0.0400)	-0.0560 (0.299)	0.0937 (0.109)
Num_Cmts	-	-	-	-	1.010*** (0.0591)	1.131*** (0.0215)
Constant	3.214*** (0.416)	2.688*** (0.313)	0.901*** (0.221)	0.736*** (0.159)	-4.605*** (1.179)	-3.006*** (0.433)
Lnalpha	-	-	-	-	0.811*** (0.137)	-0.428*** (0.0600)
N	3,187	3,187	3,187	3,187	2,557	2,557
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.193	0.172	0.128	0.127	-	-

Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

4.6.2 How Do Users Participate in Conversation Initiated by Users with Different Political Stances?

Taking advantage of users' self-disclosed political ideology in the focal subreddit, we further differentiate discourses based on creators' political stance and study how users holding different political perspectives participate in posts with various political views. Similar to our main analysis, we choose users with unclear ideology as the baseline. Some interesting results are found in Table 20. First, compared to the posts by users with unclear political stances, posts by left-leaning and right-leaning users have more engagement and attract more users to participate. Specifically, Column (1) in Table 20 shows that posts by left-leaning users receive 40% more engagement than posts by users with an unknown political stance. One possible explanation is that posts created by users with stronger and clearer political opinions are more likely to engage audiences. In contrast, it is difficult for users with ambiguous flair to identify people who share a similar ideology as them. As a result, they are less likely to receive affirmation from the community because their political identities are difficult to interpret for other community members. Thus, they may have less community attachment and a high tendency to disengage in the community (Bartel and Dutton 2001).

Interestingly, our results also imply that the comments initiated in the left-leaning and right-leaning posts contain fewer partisanship terms compared to the posts created by users that are neither left nor right-leaning. Although the estimations in the other polarization measure are not statistically significant, the sign of the estimated effects is consistent with the partisanship measure. Overall, the results imply that conversations initiated by users with clear political stances will bring more traffic to the community.

And the discussion under these posts has relatively fewer partisanship terms than posts by users with an unclear political view, who may be from the center of the political spectrum or reluctant to declare their political identity. Thus, an important practical implication to platform managers is that they need to closely monitor the discourse initiated by users with an unclear political stance, given their content can spark more partisanship in the ensuing discussion.

We further break down user engagement based on commenters' political stances and perform the above heterogeneity analysis again. We can see the detailed interaction between discourse creators and participants in Tables 21 and 22, and we find that users with different political ideologies react differently to the mandatory identity declaration. There is a significant decrease in the number of comments by users with unclear political leaning. Specifically, the total number of posts by users with unclear-leaning decreased by 66.8% after the political ideology disclosure become mandatory. For leaf-leaning and right-leaning users, we did not observe a significant change in their participation in posts created by unclear-leaning users. However, the estimated results in the interaction terms suggest that these users engage more in the left-leaning or right-leaning content. Particularly, the overall right-leaning comments and participants in posts by left-leaning users increase by 35.2% and 24.2%, respectively. These results imply that identity disclosure significantly enhances the salience of the opposite opinion and results in more interactions between users from different political sides. However, this policy also drives users without clear political leaning to leave the online discourse. In this sense, the voice from the middle ground becomes less likely to be heard when users are required to choose and declare their political identity.

Table 20. The Moderating Effect of Discourse Creator's Political Stance

	Participation				Polarization	
Variables	(1) Num_cmts	(2) Num_cmters	(3) Num_cmts_by_new_users	(4) Num_new_cmters	(6) Slang_cmts	(7) Partisanship_cmts
After	-0.215 (0.197)	-0.0306 (0.141)	-0.0122 (0.104)	0.0171 (0.0767)	0.695 (0.463)	0.479** (0.189)
Post_by_left	-0.260** (0.113)	-0.117 (0.0828)	0.00424 (0.0676)	0.0148 (0.0508)	0.513** (0.259)	0.386*** (0.117)
Post_by_right	-0.352*** (0.111)	-0.150* (0.0817)	-0.0135 (0.0684)	-0.00676 (0.0505)	0.282 (0.271)	0.196* (0.112)
After x Post_by_left	0.400** (0.168)	0.293** (0.119)	0.00259 (0.0908)	-0.0189 (0.0666)	-0.554 (0.353)	-0.468*** (0.155)
After x Post_by_Right	0.298* (0.163)	0.169 (0.116)	-0.0426 (0.0900)	-0.0398 (0.0655)	-0.453 (0.363)	-0.351** (0.151)
Discussion	1.841*** (0.0954)	1.325*** (0.0643)	0.763*** (0.0636)	0.537*** (0.0443)	0.198 (0.209)	0.0972 (0.0749)
Num_Posts_Week	-0.321*** (0.107)	-0.292*** (0.0804)	-0.0838 (0.0562)	-0.0701* (0.0403)	-0.0254 (0.302)	0.110 (0.110)
Num_Cmts	-	-	-	-	1.010*** (0.0588)	1.133*** (0.0216)
Constant	3.238*** (0.439)	2.633*** (0.329)	0.740*** (0.232)	0.590*** (0.165)	-4.912*** (1.186)	-3.221*** (0.440)
Lnalpha	-	-	-	-	0.815*** (0.137)	-0.429*** (0.0596)
N	3,187	3,187	3,187	3,187	2,557	2,557
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.194	0.173	0.126	0.125	-	-

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Regarding the right-leaning dominant group in the focal community, we do not find a significant change in slang usage after the political ideology disclosure becomes

mandatory. However, we observe increased use of partisanship terms regardless of post creators' political perspectives.

Table 21. The Interaction Between Users with Different Political Stances (Discourse Engagement)

Variables	(1) Left_Lean_Comments	(2) Left_Lean_Commenters	(3) Right_Lean_Comments	(4) Right_Lean_Comments	(5) Unclear_Lean_Comments	(6) Unclear_Lean_Commenters
After	0.0730 (0.166)	0.105 (0.109)	-0.0556 (0.175)	0.0579 (0.118)	-0.856*** (0.149)	-0.668*** (0.0894)
Post_By_Left_Lean	-0.108 (0.0938)	-0.00128 (0.0631)	-0.320*** (0.0981)	-0.153** (0.0669)	-0.426*** (0.0895)	-0.265*** (0.0610)
Post_By_Right_Lean	-0.395*** (0.0910)	-0.190*** (0.0625)	-0.217** (0.0967)	-0.0429 (0.0659)	-0.493*** (0.0894)	-0.293*** (0.0615)
After x Post_By_Left_Lean	0.279** (0.142)	0.185** (0.0914)	0.352** (0.150)	0.242** (0.0999)	0.0760 (0.128)	0.108 (0.0769)
After x Post_By_Right_Lean	0.288** (0.137)	0.176** (0.0890)	0.226 (0.146)	0.102 (0.0974)	0.0812 (0.127)	0.0964 (0.0763)
Discussion	1.448*** (0.0865)	0.952*** (0.0549)	1.695*** (0.0904)	1.149*** (0.0576)	1.205*** (0.0782)	0.777*** (0.0469)
Num_Posts_Week	-0.187** (0.0880)	-0.185*** (0.0625)	-0.272*** (0.0934)	-0.214*** (0.0668)	-0.155** (0.0667)	-0.142*** (0.0447)
Constant	1.748*** (0.347)	1.453*** (0.245)	2.480*** (0.369)	1.853*** (0.262)	2.147*** (0.273)	1.666*** (0.182)
N	3,187	3,187	3,187	3,187	3,187	3,187
Monthly fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.183	0.156	0.198	0.172	0.362	0.368

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 22. The Interaction between Users with Different Political Stances (Discourse Polarization)

Variables	(1) Slang_by_left	(2) Partisionship_by_left	(3) Slang_by_right	(4) Partisionship_by_right	(5) Slang_by_unclear	(6) Partishion_by_unclear
After	1.103*	0.825***	0.447	0.452**	-1.399*	-1.116***
	(0.569)	(0.263)	(0.586)	(0.224)	(0.785)	(0.305)
Post_By_Left_Lean	1.009***	0.584***	-0.135	0.214	-0.0628	0.210
	(0.309)	(0.168)	(0.400)	(0.147)	(0.463)	(0.179)
Post_By_Right_Lean	0.458	0.0524	0.527	0.382***	-0.421	0.252
	(0.343)	(0.160)	(0.334)	(0.137)	(0.529)	(0.164)
After x Post_By_Left_Lean	-0.793**	-0.469**	0.439	-0.150	-0.729	-0.857***
	(0.366)	(0.211)	(0.494)	(0.194)	(0.600)	(0.230)
After x Post_By_Right_Lean	-0.436	-0.177	-0.289	-0.183	-0.674	-1.048***
	(0.409)	(0.204)	(0.429)	(0.186)	(0.683)	(0.213)
Discussion	0.0656	0.141	0.222	-0.00172	0.0332	0.299
	(0.264)	(0.146)	(0.249)	(0.133)	(0.332)	(0.183)
Num_Posts_Week	-0.321	-0.0583	0.0403	0.269***	0.402	-0.0332
	(0.358)	(0.104)	(0.444)	(0.0887)	(0.471)	(0.143)
Num_Cmts	0.995***	1.101***	1.052***	1.148***	0.932***	1.084***
	(0.0715)	(0.0293)	(0.0883)	(0.0280)	(0.114)	(0.0430)
Constant	-4.765***	-4.533***	-6.274***	-3.652***	-7.214***	-5.027***
	(1.399)	(0.590)	(1.738)	(0.530)	(1.863)	(0.722)
Lnalpha	0.528**	-0.311***	0.829***	-0.462***	1.071***	-0.413**
	(0.256)	(0.0821)	(0.235)	(0.0849)	(0.386)	(0.209)
N	2,470	2,470	2,470	2,470	2,470	2,470
Monthly fixed effect	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

4.6.3 Do the Impacts Come Through the Interaction with Other Users?

Social identity literature suggests the increased polarization may come from two aspects: the increased consistent behavior between users and their social identity; and the

increased difference between the in-group and out-groups. Compared with these two factors, the inherently changed opinion should exist regardless of the presence of out-groups, whereas the latter factor should exist while interacting with other perspectives.

Table 23. The Moderating Effect of Participant View Diversity

	Participation				Polarization	
Variables	(1) Num_cmts	(2) Num_cmters	(3) Num_cmts_by_new_users	(4) Num_new_cmters	(5) Slang_cmts	(6) Partisanship_cmts
After	-0.0993 (0.140)	0.0273 (0.0913)	-0.194* (0.116)	-0.131 (0.0842)	1.132** (0.539)	0.367* (0.216)
Concentration	-2.867*** (0.109)	-2.201*** (0.0701)	-1.158*** (0.0881)	-0.878*** (0.0689)	0.527 (0.590)	-0.145 (0.268)
After x Concentration	0.0879 (0.135)	0.0925 (0.0869)	0.194* (0.105)	0.150* (0.0804)	-1.479* (0.799)	-0.334 (0.301)
Post_by_left	-0.176** (0.0694)	-0.0477 (0.0441)	0.00793 (0.0532)	0.00380 (0.0391)	0.217 (0.212)	0.152* (0.0868)
Post_by_right	-0.257*** (0.0684)	-0.0962** (0.0436)	-0.0274 (0.0523)	-0.0203 (0.0386)	0.0796 (0.214)	0.0486 (0.0856)
Discussion	1.120*** (0.0766)	0.747*** (0.0473)	0.561*** (0.0638)	0.378*** (0.0444)	0.189 (0.198)	0.0952 (0.0747)
Num_Posts_Week	0.0606 (0.0791)	0.00554 (0.0523)	0.0177 (0.0630)	0.00584 (0.0447)	-0.0553 (0.301)	0.111 (0.110)
Num_Cmts	-	-	-	-	0.978*** (0.0709)	1.100*** (0.0248)
Constant	3.942*** (0.327)	3.177*** (0.215)	1.180*** (0.261)	0.938*** (0.183)	-4.810*** (1.211)	-2.944*** (0.457)
Lnalpha	-	-	-	-	0.799*** (0.136)	-0.431*** (0.0601)
N	2,470	2,470	2,470	2,470	2,470	2,470
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.487	0.518	0.200	0.206	-	-

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Therefore, we include the interaction with different opinions into our analysis to further test the moderation effect of the interaction between different views. We measure the interaction extent between users with different political ideologies. Specifically, we utilize the Herfindahl-Hirschman Index (HHI), the widely used index of market concentration, to measure the extent to which a few political sides dominate a discourse. The higher value of HHI is, the more concentrated the discourse is, and therefore, the more likely for an echo chamber to exist. We include this measure into our analysis and get results, as shown in Table 23. The estimation of the interaction term in Column (5) suggests that when more participants are from the same side, they use less slang in their conservation. In other words, when a conversation involved users with more diverse political perspectives, more political slang is used. Despite that we do not observe the statistically significant effect in the interaction from the other two polarization measures, the sign of the estimated effect remains consistent with the slang usage.

4.7 User-level Analysis and Results

In addition to obtaining the community-level outcomes, we conduct user-level analyses to investigate how political identity disclosure affects existing users' engagement. We re-organize our data in the form of panel data recording a user's attention spent on Reddit in a week. We differentiate subreddits as focal subreddit (i.e., r/tuesday), other politics-related subreddits, and non-politics subreddits. We measure users' attention spent on each community category by the percentage of comments a user made in each category. We perform the fixed-effect analysis by adding the user fixed effect. To reveal how users with different political views react to the identity disclosure, we interact user political

stance with the key independent variable, *After*. Again, we treat users with undeclared political leaning as the baseline group.

Table 24. User Attention Allocation in Various Communities

Variables	(1) Focal_Sub	(2) Other_Politics_Subs	(3) Non_Politics_Subs
After	-0.0384*** (0.00431)	0.0354*** (0.00932)	0.00299 (0.00918)
After x Left_Lean_User	0.0320*** (0.00754)	-0.0342*** (0.0132)	0.00221 (0.0130)
After x Right_Lean_User	0.0326*** (0.00785)	-0.0475*** (0.0146)	0.0149 (0.0134)
Num_Posts_Week	-0.0238*** (0.00281)	-0.0144*** (0.00464)	-0.00945** (0.00447)
Constant	-0.0271** (0.0112)	0.553*** (0.0180)	0.474*** (0.0174)
N	23,734	23,734	23,734
User fixed effect	Yes	Yes	Yes
R-squared	0.010	0.003	0.001

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

In Table 24, we observe several results that are consistent with our observation in the community-level analyses. First, users with an undeclared political stance significantly disengage from the focal subreddit. Meanwhile, they increase their engagement in other politics-related subreddits. This result exhibits the displacement effects (Pu et al. 2020) as undeclared users switch their attention from the focal community to other neighboring ones. Similarly, left-leaning users present similar engagement patterns after implementing the identity disclosure policy, but their behavior change is at a smaller scale. In contrast, right-leaning users become less engaged in other

politics-related subreddits. Moreover, the influence of political identity disclosure does not spill over to non-politics subreddits. Users' attention spent in non-politics subreddits does not significantly change regardless of their political leaning.

Overall, our results on users' attention spent validate our community-level findings and show that identity disclosure policy will disapprovingly affect users with different political stances. It attracts users who share the views with most community members but discourages users with minor views.

4.8 Conclusion and General Discussion

4.8.1 Theoretical Implication

Our research contributes to three streams of research. First, our work contributes to the identity disclosure literature in the Information Systems (Cavusoglu et al. 2016. Lu et al. 2019, Pu et al. 2020). Prior literature has investigated identity disclosure, but these papers mainly focus on users' personal identity. Our study extends this stream of research by studying an understudied yet important social identity, user political stance. Second, in addition to user engagement, our research further investigates the impacts of identity disclosure on polarization in online discourses. Our results suggest that political identity disclosure may improve idea exchange, but meanwhile, it can also lead to more polarized content and more partisanship in the interaction.

Second, our work contributes to the growing political polarization literature, especially the studies (Bail et al. 2018; Greenstein et al. 2016). Unlike extant literature that mainly concentrates on the consequences of inter-group interaction and information exposure, our research focuses on political identity salience in the online discourse. We

add empirical evidence showing that online discourses are more likely to get intense under the strengthened distinction between in-group and out-groups. Notably, a significant increase iin polarization is found in users with minor political views in the community.

Third, our work contributes to the growing online content moderation literature (Jaidka et al. 2019; Matias 2019 b). Prior literature is relatively silent on identity disclosure policy in community management. Our study fills this gap by showing that presenting users' political stances will stimulate conversations between users with different political perspectives. Meanwhile, it also results in more polarized interaction. Further, we demonstrate that the increased partisanship is contingent on content creators' political stance. Therefore, platform moderators should prioritize their moderation on discourses initiated by users from unclear political positions as these content usually trigger more polarization.

4.8.2 Practical Implication

Our results also shed light on the practitioners of social media platforms. Our work highlights the trade-off of identity declaration. Results suggest that displaying users' political ideology will increase the interaction between users with different opinions. However, such increased interaction also results in more polarized content and partisanship in the platform. Therefore, platform managers must be cautious about implementing the identity disclosure policy based on their platform growth stage and moderation capacity. For platforms with the goal of opinion exchange, displaying user political stances might be beneficial because it would encourage more intergroup

interactions. Meanwhile, managers should monitor such interactions more carefully to avoid the potentially rising polarization.

Our study further suggests that mandatory identity disclosure may result in structural changes in their user base. Such policy will decrease users with undecided political perspectives or unwilling to declare their political stances. For the whole platform managers, this may lead to unexpected consequences because it discourages users in the center position from participating. Thus, the online discourse will lose the voice of the middle ground. To retain users who are unsure or who do not want to show their political side, platform practitioners may need to consider optimizing the ideology category design by offering more suitable options for their users. For example, instead of forcing users to either side of the political spectrum, some communities allow users to choose undecided or customize their flair to describe their political stance. The diversity and flexibility of political stance would retain more users in the community. It could also ease the salience of in-group and out-group and ease the tension in online political discourse. Moreover, platform managers should pay more attention to participants with minority opinions because their engagement is likely to become more partisan in the identity-declared environment. Political stance disclosure will help communities to attract more users who share the same political perspective as the majority. Meanwhile, it also amplifies the political discussion by retaining users who hold stronger and opposite opinions. As a result, the disclosed political identities may hurt the conversation harmony in the community.

Lastly, in terms of content moderation priority, platform managers should prioritize their moderation to discussion-centered conversations, given that discussion can

engage more users (newcomers in particular). Meanwhile, compared to user-initiated conversations, content moderators should primarily pay attention to content posted by users with unclear political stances because such content tends to be associated with more polarized content.

4.8.3 Limitation and Future Work

We are aware that our work is not without limitations. First, the studied subreddit is dominated by the center-right group. Therefore, the left-leaning users are more like the guest and minority group in this context. Considering the mentality and behavioral differences between the left-leaning and right-leaning individuals (Bail et al. 2018; Frimer et al. 2017; Graham et al. 2009; Jost et al. 2007), our results may not generalize to contexts where users with left-leaning political views dominate the online discourse. To advance this research forward, researchers can extend this study to political discussion contexts with various combinations of participants to further investigate whether the role of political ideology is also contingent on the place of the dominant group's political stance.

Second, in our study, we label users with unclear or none flair as the unclear-leaning group. Theoretically, this group may mix users with a neutral political stance, users who reluctant to reveal their actual identity, and users who are unsure about their political identity. However, we cannot differentiate these two types of users in our study due to the data availability. In the future, researchers can consider recruiting users from these three minor groups and then perform lab experiments to test the impact of political identity disclosure on these three minor groups.

Third, in our research context, the identity disclosure policy may discourage users from participating in the community through two aspects. One is the enhanced entry barrier because users need to set up their flair manually before they join the discussion. The other aspect is the identity-related mechanism that we discuss in the research. We try to disentangle the impact of the extra participation cost from the identity disclosure by comparing the behavior changes between users who declared their identity before the policy change and users who did not. The rationale behind this is that users who had already disclosed their identity would not experience the extra participation cost after the policy change. However, they still experience environmental change through the declaration by other users. The results in Appendix E indicate that the effect of political identity disclosure persists for users without experiencing extra participation costs but at a smaller scale. Meanwhile, we are also aware that these results only reveal the impacts of the community norm change but do not completely differentiate the influence of generic identity disclosure and political one. In the future, researchers can consider employing experiments to unpack the impacts of identity declaration.

Last but not least, our work currently relies on the declared user identity. However, it is still possible that some users misuse this feature and being dishonest about their true identity. This is also the primary concern of human moderators in the focal subreddit. To eliminate this concern in our analysis, we take advantage of moderators' continuing effort in monitoring the inconsistency between users' declared identity and their activities. We collect data across three years and only include users who had consistent flair in the longer period. More than 95% of users remained after this step. Moving forward, we plan to further reduce this flair misuse concern by measuring the

consistency between the declared identity and user comments. Also, given that the focus of political discourses and the polarized term may change over time (Greenstein et al. 2016; Gentzkow et al. 2010; Gentzkow et al. 2019), we can apply alternative text-based polarization measures to validate our results further.

CHAPTER 5

CONCLUSION

User engagement is the main driving force for online platform growth. With the rapidly changing technological and societal environment, online platforms take various approaches to motivate user engagement and achieve a better online environment. In this dissertation, with data from three online platforms, I conduct empirical analyses to examine the platform policies and their impacts on user engagement.

In Study 1, I conduct my research on the goal-pursuit platforms. Motivated by that extant studies are silent on the interaction between technology adoption and goal pursuit, I fill this research gap by empirically investigating multi-channel adoption's impact on users' goal pursuit, particularly goal pursuit effort and persistence. Viewing mobile adoption as a natural treatment, I conduct our research on Picmonic, an Exam Prep platform in the U.S. With the estimations of PSM-based DiD and several robustness checks, the results suggest that multi-channel adoption increases the overall users' goal pursuit effort by 140.1%. Such a positive impact on goal pursuit persistence is also observed. Adoption of the mobile channel also leads to the diversity of goal pursuit activities. Interestingly, more substantial motivational effects have been found on users with a specific goal and higher goal pursuit competency. Overall, I conclude that strategic channel extension and user intervention are necessary to better assist users' goal pursuit.

In Study 2, I turn my attention to the rising group of online participants, volunteer moderators, and study the influences of machine-powered regulations on their engagement. I collect moderation records from Reddit and investigate the impact of

machine-powered governance on volunteer human moderation. With data collected from 156 subreddits, I found that delegating moderation to machines augments volunteer moderators' role as community managers. Human moderators present more moderation-related engagement, including both corrective and supportive interactions with their community members. Notably, the results indicate that such effects manifest among communities with large user bases and detailed community guidelines, suggesting that community needs for moderation is the driving factor for volunteer moderators' increased contributions.

Lastly, in Study 3, I focus on identity declaration and its influence on user engagement and polarization in subsequent political discourses. Taking advantage of a community policy change on Reddit, I find that when individual political identity becomes more transparent and salient in online discourses, the interaction between users with opposing views increases. However, at the same time, such interaction becomes more polarized. Notably, the left-leaning users, the minority group, maintain a similar level of engagement as before, but they use more slang and partisanship terms in their subsequent discourses. In contrast, another minority group, users with ambiguous political identity turn their attention to other politics-related communities and disengage from the focal community.

REFERENCES

Alfes, K., Shantz, A., & Bailey, C. (2016). Enhancing volunteer engagement to achieve desirable outcomes: What can non-profit employers do?. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 27(2), 595-617.

An, J., Kwak, H., Posegga, O., & Jungherr, A. (2019). Political discussions in homogeneous and cross-cutting communication spaces. *Proceedings of the International AAAI Conference on Web and Social Media*, 13, 68-79.

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Anzalone, J. (2020). *Characteristics that motivate a volunteer workforce: A case study of one of america's largest volunteer organizations* [Doctoral dissertation]. University of Nebraska - Lincoln.

Autor, D. H. & Dorn, D., (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review*, 103(5), 1553-97.

Babar, Y., & Burtch, G. (2020). Examining the heterogeneous impact of ride-hailing services on public transit use. *Information Systems Research*, 31(3), 820-834.

Bai, B., Dai, H., Zhang, D., Zhang, F., & Hu, H. (2020, April 2). The impacts of algorithmic work assignment on fairness perceptions and productivity: Evidence from field experiments. *Available at SSRN*. https://dx.doi.org/10.2139/ssrn.3550887

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216-9221.

Bang, Y., Lee, D., Han, K., Hwang, M., & Ahn, J. (2013 a). Channel capabilities, product characteristics, and the impacts of mobile channel introduction. *Journal of Management Information Systems*, 30(2), 101–126.

Bang, H., Ross, S., & Reio, T. G. (2013 b). From motivation to organizational commitment of volunteers in non-profit sport organizations. *Journal of Management Development*, 32(1), 96-112.

Bapna R, Ramaprasad J, & Umyarov A. (2018). Monetizing freemium communities: Does paying for premium increase social engagement?. *MIS Quarterly*, 42(3), 719–735.

Bartel, C., & Dutton, J. (2001). Ambiguous organizational memberships: Constructing organizational identities. *Social Identity Processes in Organizational Contexts*, 115-130.

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, 24(2), 117–150.

Birman, I. (2018). *Moderation in different communities on Reddit--A qualitative analysis study*. Georgia Institute of Technology.

Blumenfeld, P. C. (1992). Classroom learning and motivation: Clarifying and expanding goal theory. *Journal of Educational Psychology,* 84(3), 272.

Brandstätter, V., & Frank, E. (2002). Effects of deliberative and implemental mindsets on persistence in goal-directed behavior. *Personality and Social Psychology Bulletin,* 28(10), 1366-1378.

Brewer, Marilynn, & Rupert J. Brown. (1998). Intergroup Relations. *In The Handbook of Social Psychology*, 4th ed., vol. 2, eds. D. T. Gilbert, S. T. Fiske, and G. Lindzey. Boston: McGraw-Hill, 554–594.

Burtch, G., Ghose, A., & Wattal, S. (2016). Secret admirers: An empirical examination of information hiding and contribution dynamics in online crowdfunding. *Information Systems Research,* 27(3), 478–496.

Burtch, G., Carnahan, S., & Greenwood, B. N. (2018). Can you hig it? An empirical examination of the gig economy and entrepreneurial activity. *Management Science*, 64(12), 5497-5520.

Burtch, G., He, Q., Hong, Y., & Lee, D. (2020). Peer recognition increases user content generation but reduces content novelty. *Proceeding of the 40th International Conference on Information Systems.*

Butler, B. S., Bateman, P. J., Gray, P. H., & Diamant, E. I. (2014). An attraction–selection–attrition theory of online community size and resilience. *MIS Quarterly*, 38(3), 699-729.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys,* 22(1), 31-72.

Campion, M. A., & Lord, R. G. (1982). A control systems conceptualization of the goal-setting and changing process. *Organizational Bbehavior and Human Performance,* 30(2), 265-287.

Catherine B., & Soraya C. (2016, April 13). The secret rules of the Internet. *The Verge*. http://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech

Cavusoglu, H., Phan, T. Q., Cavusoglu, H., & Airoldi, E. M. (2016). Assessing the impact of granular privacy controls on content sharing and disclosure on Facebook. *Information Systems Research*, 27(4), 848-879.

Chae, M., & Kim, J. (2004). Size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. *Behaviour and Information Technology,* 23(3), 165–81.

Chandrasekharan, E., Mattia S., Jhaver S., Charvat H., Bruckman A., Lampe C., Eisenstein J., & Gilbert E. (2018). The Internet's hidden rules: An empirical study of Reddit norm violations at micro, meso, and macro scales. *Proceedings of the ACM on Human-Computer Interaction*, 2 (CSCW). https://doi.org/10.1145/3274301

Chandrasekharan, E., & Gilbert, E. (2019, July 17). Hybrid approaches to detect comments violating macro norms on Reddit. *arXiv preprint arXiv:1904.03596*. https://arxiv.org/pdf/1904.03596.pdf

De Leon, F. L. L., & Rizzi, R. (2016). Does forced voting result in political polarization?. *Public Choice*, 166(1-2), 143-160.

Demszky, D., Garg, N., Voigt, R., Zou, J., Gentzkow, M., Shapiro, J., & Jurafsky, D. (2019, April 4). Analyzing polarization in social media: Method and application to Tweets on 21 mass shootings. *arXiv preprint arXiv:1904.01596*. https://arxiv.org/pdf/1904.01596.pdf

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018, May 24). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.* https://arxiv.org/pdf/1810.04805.pdf&usg=ALkJrhhzxlCL6yTht2BRmH9atgvKFxHsxQ

Dimock, M., Doherty, C., Kiley, J., & Oates, R. (2014). Political polarization in the American public. Pew Research Center, 12.

Dixon, J., Hong, B., & Wu, L. (2020, June 3). The robot revolution: Managerial and employment consequences for firms. NYU Stern School of Business. *Available at SSRN*. https://dx.doi.org/10.2139/ssrn.3422581

Donnelly Jr, J. H., & Ivancevich, J. M. (1975). Role clarity and the salesman: An empirical study reveals that perceived role clarity may be an important factor in maximizing a salesman's job performance. *Journal of Marketing*, 39(1), 71-74.

Dosono, B., & Semaan, B. (2019). Moderation practices as emotional labor in sustaining online communities: The case of AAPI identity work on Reddit. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290605.3300372

Feather, N. T. (1962). The study of persistence. *Psychological Bulletin,* 59(2), 94.

Fiesler, C., Jiang, J., McCann, J., Frye, K., & Brubaker, J. R. (2018). Reddit rules! Characterizing an ecosystem of governance. *Proceedings of the Twelfth International AAAI Conference on Web and Social Media*. https://ojs.aaai.org/index.php/ICWSM/article/download/15033/14883

Fishbach, A., & Dhar, R. (2005). Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *Journal of Consumer Research,* 32(3), 370–377.

Fishbach, A., & Ferguson, M. J. (2007). The goal construct in social psychology. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (p. 490–515). The Guilford Press.

Fishbach, A., & Finkelstein, S. R. (2012). How feedback influences persistence, disengagement, and change in goal pursuit. *Goal-directed Behavior*, 203-230. Psychology Press.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19(3), 291-313.

Fowler, J. H., & Kam, C. D. (2007). Beyond the self: Social identity, altruism, and political participation. *The Journal of Politics*, 69(3), 813-827.

Frimer, J. A., Skitka, L. J., & Motyl, M. (2017). Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *Journal of Experimental Social Psychology*, (72), 1-12.

Garg, R., & Telang, R. (2018). To be or not to be linked: Online social networks and job search by unemployed workforce. *Management Science*, 64(8), 3926-3941.

Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, 78(1), 35-71.

Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, 87(4), 1307-1340.

Gerber, A. S., Huber, G. A., Doherty, D., & Dowling, C. M. (2012). Personality and the strength and direction of partisan identification. *Political Behavior*, 34(4), 653-688.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029.

Greene, S. (2004). Social identity theory and party identification. *Social Science Quarterly*, 85(1), 136-153.

Greenstein, S., Gu, Y., & Zhu, F. (2016). Ideological segregation among online collaborators: Evidence from Wikipedians. National Bureau of Economic Research, No. w22744.

Greenwood, B. N., & Wattal, S. (2017). Show me the way to go home: An empirical investigation of ride-sharing and alcohol related motor vehicle fatalities. *MIS Quarterly*, 41(1), 163-187.

Grimmelmann, J. (2015). *The virtues of moderation*. Yale JL & Tech., 17, 42.

Gollatz, K., Beer, F., & Katzenbach, C. (2018). The turn to artificial intelligence in governing communication online. *Social Science Open Access Repository*. https://www.ssoar.info/ssoar/handle/document/59528

Hammer, H. L. (2016). Automatic detection of hateful comments in online discussion. In: Maglaras L., Janicke H., Jones K. (eds) *Industrial Networks and Intelligent Systems. INISCOM 2016. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. https://doi.org/10.1007/978-3-319-52569-3_15

Han, X., and Hu, M. M. Intensified ideological online clashes with group political bias. https://www.tse-fr.eu/sites/default/files/TSE/documents/conf/2021/han.pdf

Hassan, S. (2013). The importance of role clarification in workgroups: Effects on perceived role clarity, work satisfaction, and turnover rates. *Public Administration Review*, 73(5), 716-725.

Higgins, E. T. (2000). Making a good decision: Value from fit. *American Psychologist,* 55(11), 1217–1230

Higgins, E. T., Idson, L. C., Freitas, A. L., Spiegel, S., & Molden, D. C. (2003). Transfer of value from fit. *Journal of Personality and Social Psychology,* 84(6), 1140–1153.

Huang, N., Zhang, J., Burtch, G., Li, X., & Chen, P. (2021). Combating procrastination on massive online open courses via optimal calls to action. *Information Systems Research*, Forthcoming.

Huang, S. C., & Zhang, Y. (2011). Motivational consequences of perceived velocity in consumer goal pursuit. *Journal of Marketing Research,* 48(6), 1045–1056.

Huang, L., Lu X., & Ba, S. (2016). An empirical study of the cross-channel effects between web and mobile shopping channels. *Information and Managemen*t, 53(2), 265–278.

Jaidka, K., Zhou, A., & Lelkes, Y. (2019). Brevity is the soul of Twitter: The constraint affordance and political discussion. *Journal of Communication*, 69(4), 345-372.

Jhaver, S., Vora, P., & Bruckman, A. (2017). Designing for civil conversations: Lessons learned from ChangeMyView. *GVU Technical Report*. https://smartech.gatech.edu/bitstream/handle/1853/59080/cmv_chi_paper.pdf?sequence=1&isAllowed=y

Jhaver, S., Birman, I., Gilbert, E., & Bruckman, A. (2019 a). Human-machine collaboration for content regulation: The case of Reddit automoderator. *ACM Transactions on Computer-Human Interaction*, 26(5), 1-35. https://doi.org/10.1145/3338243

Jhaver, S., Appling, D. S., Gilbert, E., & Bruckman, A. (2019 b). Did you suspect the post would be removed? Understanding user reactions to content removals on Reddit. *Proceedings of the ACM on Human-Computer Interaction, 3(CSCW)*. https://doi.org/10.1145/3359294

Jhaver, S., Bruckman, A., & Gilbert, E. (2019 c). Does transparency in moderation really matter? User behavior after content removal explanations on Reddit. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW). https://doi.org/10.1145/3359252

Jiao, J. J., & Cole, C. (2015). The effects of goal publicity on goal persistence in the social media world. *Consumer Psychology in a Social Media World*, 139-160. Routledge.

Jost, J. T., Napier, J. L., Thorisdottir, H., Gosling, S. D., Palfai, T. P., & Ostafin, B. (2007). Are needs to manage uncertainty and threat associated with political conservatism or ideological extremity?. *Personality and Social Psychology Bulletin*, 33(7), 989-1007.

Jung, J., Bapna R., Ramaprasad J., & Umyarov, A. (2019). Love unshackled: Identifying the effect of mobile app adoption in online dating. *MIS Quarterly,* 43(1), 47-72.

Karusala, N., Vishwanath, A., Kumar, A., Mangal, A., & Kumar, N. (2017). Care as a resource in underserved learning environments. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW). https://doi.org/10.1145/3134739

Kilner P. G., Hoadley C. M. (2005). Anonymity options and professional participation in an online community of practice. *Proceedings of 2005 Conference on Computer Support for Collaborative Learning: Learning: The Next 10 Years!*

Klein, H. J., Whitener, E. M., & Ilgen, D. R. (1990). The role of goal specificity in the goal-setting process. *Motivation and Emotion,* 14(3), 179-193.

Kruglanski, A. W., Pierro, A., & Sheveland, A. (2011). How many roads lead to rome?

Equifinality set-size and commitment to goals and means. *European Journal of Social Psychology,* 41(3), 344–352.

Kumar, N., Qiu, L., & Kumar, S. (2018). Exit, voice, and response on digital platforms: An empirical investigation of online management response strategies. *Information Systems Research,* 29(4), 849-870.

Kummer, M., Slivko, O., & Zhang, X. (2020). Unemployment and digital public goods contribution. *Information Systems Research*, 31(3), 801-819.

Künsting, J., Wirth, J., & Paas, F. (2011). The goal specificity effect on strategy use and instructional efficiency during computer-based scientific discovery learning. *Computers and Education,* 56(3), 668-679.

Lacus S. M., King G., & Porro G. (2009). CEM: Software for coarsened exact matching. *Journal of Statistical Software,* 30(13), 1-27.

Lee, Z., Chan, T., Chong, A., & Thadani, D. (2017). An empirical investigation into the antecedents and consequences of customer engagement in omnichannel retailing. *Proceeding of the 21st Pacific Asia Conference on Information Systems*.

Leonardi, P. M. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly,* 35(1), 147–168.

Levy, R. E. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review*, 111(3), 831-70.

Li, Z., Fang, X., Bai, X., & Sheng, O. R. L. (2017). Utility-based link recommendation for online social networks. *Management Science*, 63(6), 1938-1952.

Liu, J., Abhishek, V., & Li, B. (2016). The impact of mobile technology on customer behavior. *Proceedings of the 37th International Conference on Information Systems*.

Locke, E. A. (1996). Motivation through conscious goal setting. *Applied and Preventive Psychology,* 5(2), 117-124.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin,* 90(1), 125.

Long, K., Vines, J., Sutton, S., Brooker, P., Feltwell, T., Kirman, B., Barnett, J., & Lawson, S. (2017). Could you define that in bot terms? Requesting, creating and using bots on Reddit. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3488-3500. http://dx.doi.org/10.1145/3025453.3025830

Lu, Y., Gupta, A., Ketter, W., & Van Heck, E. (2019). Information transparency in business-to-business auction markets: The role of winner identity

disclosure. *Management Science*, 65(9), 4261-4279.

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, 38(6), 937-947.

Lyons, T. F. (1971). Role clarity, need for clarity, satisfaction, tension, and withdrawal. *Organizational Behavior and Human Performance*, 6(1), 99-110.

Markman, A. B., Brendl, C. M., & Kim, K. (2007). Preference and the specificity of goals. *Emotion,* 7(3), 680.

Matias, J. N. (2019 a). The civic labor of volunteer moderators online. *Social Media+ Society*. https://journals.sagepub.com/doi/pdf/10.1177/2056305119836778

Matias, J. N. (2019 b). Preventing harassment and increasing group participation through social norms in 2,190 online science discussions. *Proceedings of the National Academy of Sciences*, 116(20), 9785-9789.

Menking, A., & Erickson, I. (2015). The heart work of Wikipedia: Gendered, emotional labor in the world's largest online encyclopedia. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. http://dx.doi.org/10.1145/2702123.2702514

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303-330.

Orr, G. (2010). A review of literature in mobile learning: Affordances and constraints. *The 6th IEEE International Conference on Wireless, Mobile and Ubiquitous Technologies in Education*. doi: 10.1109/WMUTE.2010.20.

Oyserman, D., & Dawson, A. (2020). Your fake news, our facts: Identity-based motivation shapes what we believe, share, and accept. In Greifeneder, R., Jaffé, M., Newman, E.J., & Schwarz, N. (Eds.) *The Psychology of Fake News: Accepting, Sharing, and Correcting Misinformation*. London, UK: Psychology Press.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751.

Petty, R. E., Wegener, D. T., & White, P. H. (1998). Flexible correction processes in social judgment: Implications for persuasion. *Social Cognition*, 16(1), 93-113.

Pu, J., Chen, Y., Qiu, L., & Cheng, H. K. (2020). Does identity disclosure help or hurt user content generation? Social presence, inhibition, and displacement effects. *Information Systems Research*, 31(2), 297-322.

Redding, S. (2014). Personal competency: A framework for building students' capacity to

learn, Center on Innovations in Learning, Temple University https://files.eric.ed.gov/fulltext/ED558070.pdf .

Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies*, 28(3), 377-408.

Ren, Y., Harper, F. M., Drenner, S., Terveen, L., Kiesler, S., Riedl, J., & Kraut, R. E. (2012). Building member attachment in online communities: Applying theories of group identity and interpersonal bonds. *MIS Quarterly*, 36(3), 841-864.

Ren, Y., & Kraut, R. E. (2014). Agent-based modeling to inform online community design: Impact of topical breadth, message volume, and discussion moderation on member commitment and contribution. *Human–Computer Interaction*, 29(4), 351-389.

Riediger, M., & Freund, A. M. (2004). Interference and facilitation among personal goals: Differential associations with subjective well-being and persistent goal pursuit. *Personality and Social Psychology Bulletin,* 30(12), 1511-1523.

Roberts, S. T. (2014). Behind the screen: *The hidden digital labor of commercial content moderation* [Doctoral dissertation]. University of Illinois at Urbana-Champaign.

Rogowski, J. C., & Sutherland, J. L. (2016). How ideology fuels affective polarization. *Political Behavior*, 38(2), 485-508.

Ruckenstein, M., & Turunen, L. L. M. (2019). Re-humanizing the platform: Content moderators and the logic of care. *New Media & Society*, 22(6), 1026-1042.

Rubin, D. (2006). *Matched sampling for causal effects*. Cambridge University Press.

Santhanam, R., Liu, D., & Shen, W. C. M. (2016). Research note—Gamification of technology-mediated training: Not all competitions are the same. *Information Systems Research*, 27(2), 453-465.

Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, 26(3), 207-231.

Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading andWriting Quarterly,* 19(2), 159-172.

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of 'humanizing customer service chatbots. *Information Systems Research,* Forthcoming.

Seering, J., Kraut, R., & Dabbish, L. (2017). Shaping pro and anti-social behavior on Twitch through moderation and example-setting. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 111-125.

Seering, J., Wang, T., Yoon, J., & Kaufman, G. (2019). Moderator engagement and community development in the age of algorithms. *New Media & Society*, 21(7), 1417-1443.

Sitzmann, T., & Yeo, G. (2013). A meta-analytic investigation of the within-person self-efficacy domain: Is self-efficacy a product of past performance or a driver of future performance?. *Personnel Psychology,* 66(3), 531-568.

Smith, D. H. (1994). Determinants of voluntary association participation and volunteering: A literature review. *Nonprofit and Voluntary Sector Quarterly*, 23(3), 243-263.

Smith, K. G., Locke, E. A., & Barry, D. (1990). Goal setting, planning, and organizational performance: An experimental simulation. *Organizational Behavior and Human Decision Processes,* 46(1), 118-134.

Srinivasan, K. B., Danescu-Niculescu-Mizil, C., Lee, L., and Tan, C. (2019). Content removal as a moderation strategy: Compliance and other outcomes in the ChangeMyView community. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW). https://doi.org/10.1145/3359265

Son, Y., Oh, W., Han, S.P., & Park, S. (2016). The adoption and use of mobile application- based reward systems : Implications for offline purchase and mobile commerce. *Proceedings of the 37th International Conference on Information Systems*.

Swann, C., and Rosenbaum S. (2018). Do we need to reconsider best practice in goal setting for physical activity promotion?. *British Journal of Sports Medicine,* 52(8), 485-486.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3), 755-769.

Tan, T. F., & Netessine, S. (2020). At your service on the table: Impact of tabletop technology on restaurant performance. *Management Science*, 66(10), 4496-4515.

Tubbs, M. E. (1986). Goal setting: A meta-analytic examination of the empirical evidence. *Journal of Applied Psychology,* 71(3), 474.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the Social Group: A self-categorization Theory*. Basil Blackwell.

Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin*, 20(5), 454-463.

Uetake, K., & Yang, N. (2018, March 7). Harnessing the small victories: Goal design strategies for a mobile calorie and weight loss tracking application. *Available at SSRN 2928441*. https://dx.doi.org/10.2139/ssrn.2928441

Urman, A. (2020). Context matters: Political polarization on Twitter from a comparative perspective. *Media, Culture & Society*, 42(6), 857-879.

Van Alstyne, M. W., Parker, G. G., & Choudary, S. P. (2016). Pipelines, platforms, and the new rules of strategy. *Harvard Business Review*, 94(4), 54-62.

Vecina, M. L., Chacón, F., Marzana, D., & Marta, E. (2013). Volunteer engagement and organizational commitment in nonprofit organizations: What makes volunteers remain within organizations and feel happy?. *Journal of Community Psychology*, 41(3), 291-302.

Wallace, S. G., & Etkin, J. (2018). How goal specificity shapes motivation: A reference points perspective. *Journal of Consumer Research,* 44(5), 1033-1051.

Williams, M. (2007). Policing and cybersociety: The maturation of regulation within an online community. *Policing & Society*, 17(1), 59-82.

Wooldridge J. (2002). *Econometric analysis of cross section and panel data*. MIT Press, Cambridge, MA.

Xu, K., Chan, J., Ghose, A., & Han, S.P. (2017). Battle of the channels: The impact of tablets on digital commerce. *Management Science,* 63(5), 1469–1492.

Yang, Y. (2019). When power goes wild online: How did a voluntary moderator's abuse of power affect an online community?. *Proceedings of the Association for Information Science and Technology*, 56(1), 504-508.

Yu, B., Seering, J., Spiel, K., & Watts, L. (2020). Taking care of a fruit tree: Nurturing as a layer of concern in online community moderation. *In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. http://dx.doi.org/10.1145/3334480.3383009

Yu, B., Spiel, K., & Watts, L. (2018). Supporting care as a layer of concern: Nurturing attitudes in online community moderation. *In Sociotechnical Systems of Care: A CSCW18 Workshop.*

Zhang, P. (2008). Motivational affordances: Resaons for ICT design and use. *Communications of the ACM,* 51(11), 145–147.

Zheng, L., Albano, C. M., Vora, N. M., Mai, F., & Nickerson, J. V. (2019). The roles bots play in Wikipedia. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW). https://doi.org/10.1145/3359317

APPENDIX A

ROBUSTNESS CHECK FOR THE IMPACT OF MODEL ADOPTION

Table 25. Main Estimation Results Using Relative Adoption Time

	Num_Card		Card_Day		Num_Quiz		Quiz_Day	
	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC
T-2	0.059 (0.083)	0.059 (0.083)	0.018 (0.038)	0.018 (0.038)	0.046 (0.071)	0.046 (0.071)	0.021 (0.036)	0.021 (0.036)
T+0	1.946*** (0.086)	1.162*** (0.089)	0.861*** (0.037)	0.446*** (0.038)	1.501*** (0.074)	0.886*** (0.074)	0.733*** (0.037)	0.404*** (0.037)
T+1	1.346*** (0.095)	0.934*** (0.092)	0.645*** (0.044)	0.422*** (0.042)	1.101*** (0.082)	0.721*** (0.078)	0.582*** (0.042)	0.365*** (0.040)
T+2	1.002*** (0.097)	0.695*** (0.095)	0.491*** (0.044)	0.322*** (0.042)	0.810*** (0.081)	0.535*** (0.079)	0.428*** (0.042)	0.272*** (0.040)
No. of Obs.	7,620	7,620	7,620	7,620	7,620	7,620	7,620	7,620
R-Squared	0.446	0.419	0.448	0.422	0.402	0.375	0.402	0.375

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User control and time fixed effect are included.

Table 26. Main Estimation Results Using Fixed Effect

	Num_Card		Card_Day		Num_Quiz		Quiz_Day	
	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC	PC+Mobi	PC
$After_{it}$	1.176*** (0.073)	0.456*** (0.075)	0.522*** (0.031)	0.140*** (0.033)	0.953*** (0.063)	0.358*** (0.062)	0.451*** (0.031)	0.129*** (0.031)
$Mobile_i \times After_{it}$	1.436*** (0.074)	0.913*** (0.074)	0.670*** (0.033)	0.389*** (0.033)	1.145*** (0.063)	0.704*** (0.062)	0.583*** (0.032)	0.340*** (0.031)
No. of Obs.	7,620	7,620	7,620	7,620	7,620	7,620	7,620	7,620
Number of user_id_fe	1,524	1,524	1,524	1,524	1,524	1,524	1,524	1,524

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User fixed and time fixed effect are included.

Table 27. Main Estimation Results Using LA-PSM

	Num_Card		Card_Day		Num_Quiz		Quiz_Day	
	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC
$Mobile_i$	-0.007 (0.057)	-0.007 (0.056)	-0.015 (0.107)	-0.016 (0.106)	0.011 (0.049)	0.008 (0.048)	0.072 (0.096)	0.011 (0.096)
$After_{it}$	-0.487*** (0.072)	-0.492*** (0.072)	-0.690*** (0.109)	-0.700*** (0.109)	-0.372*** (0.058)	-0.372*** (0.058)	-0.562*** (0.097)	-0.528*** (0.095)
$Mobile_i \times After_{it}$	1.138*** (0.108)	0.660*** (0.107)	1.656*** (0.168)	0.904*** (0.160)	0.851*** (0.091)	0.438*** (0.088)	1.308*** (0.157)	0.180 (0.131)
No. of Obs.	3,890	3,890	3,890	3,890	3,890	3,890	3,890	3,890
R-Squared	0.389	0.381	0.359	0.367	0.372	0.358	0.321	0.324

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User control and time fixed effect are included.

Table 28. Estimation Results Using CEM

	Num_Card		Card_Day		Num_Quiz		Quiz_Day	
	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC
$Mobile_i$	-0.063 (0.076)	-0.060 (0.074)	-0.009 (0.104)	-0.012 (0.102)	-0.026 (0.063)	-0.045 (0.060)	-0.004 (0.100)	-0.136 (0.092)
$After_{it}$	-0.316*** (0.093)	-0.313*** (0.093)	-0.239* (0.109)	-0.243* (0.108)	-0.250** (0.079)	-0.244** (0.079)	-0.268* (0.109)	-0.277* (0.108)
$Mobile_i \times After_{it}$	1.404*** (0.143)	0.833*** (0.144)	1.618*** (0.185)	0.837*** (0.172)	1.128*** (0.124)	0.652*** (0.120)	1.480*** (0.185)	0.518*** (0.149)
No. of Obs.	1,560	1,560	1,560	1,560	1,560	1,560	1,560	1,560
R-Squared	0.374	0.322	0.339	0.316	0.347	0.295	0.306	0.255

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User control and time fixed effect are included.

Table 29. Falsification Test

	Num_Card	Card_Day	Num_Quiz	Quiz_Day
$Mobile_i$	0.039 (0.069)	0.017 (0.031)	0.043 (0.057)	0.029 (0.029)
$After_{it}$	0.401*** (0.055)	0.175*** (0.026)	0.303*** (0.046)	0.156*** (0.024)
$Mobile_i \times After_{it}$	-0.052 (0.083)	-0.015 (0.038)	-0.040 (0.071)	-0.020 (0.036)
No. of Obs.	3,048	3,048	3,048	3,048
R-Squared	0.669	0.649	0.628	0.618

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User control and time fixed effect are included.

APPENDIX B

VADLIDATE RESULTS WITH NURSING STUDENT DATA

Table 30. T-test Results After Matching

N	Variable	Mean (Control)	Mean (Treated)	*t*-value	*p*-value
1,036	$Tenure\_Wk_i$	15.313 (0.670)	14.896 (0.628)	0.454	0.650
1,036	$Has\_Playlist_i$	0.378 (0.021)	0.378 (0.021)	0.000	1.000
1,036	$Num\_Card\_Pre_i$	2.847 (0.073)	2.905 (0.074)	-0.560	0.575
1,036	$Num\_Quiz\_Pre_i$	2.279 (0.065)	2.323 (0.067)	-0.474	0.636
1,036	$Card\_Day\_Pre_i$	1.470 (0.040)	1.502 (0.038)	-0.576	0.565
1,036	$Quiz\_Day\_Pre_i$	1.328 (0.039)	1.352 (0.037)	-0.441	0.660
1,036	$Num\_Card\_2w_i$	0.936 (0.060)	0.981 (0.064)	-0.517	0.605
1,036	Cur_paid_i	0.846 (0.016)	0.844 (0.016)	0.086	0.932

Notes: (1) Standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card_Pre, Num_Quiz_Pre*$_i$, *Card_Day_Pre*$_i$ and *Quiz_Day_Pre*$_i$ are log transformed; (3) Caliper of 0.05 is used to generate the matched pairs.

Table 31. Main Estimation Results of Users' Goal Pursuit on Overall and PC Channel

	Num_Card		Card_Day		Num_Quiz		Quiz_Day	
	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC
$Mobile_i$	0.044 (0.045)	0.041 (0.042)	0.008 (0.020)	0.008 (0.020)	0.020 (0.038)	0.008 (0.037)	-0.007 (0.019)	-0.006 (0.019)
$After_{it}$	-0.184*** (0.051)	-0.192*** (0.051)	-0.106*** (0.022)	-0.106*** (0.022)	-0.185*** (0.041)	-0.194*** (0.041)	-0.125*** (0.021)	-0.125*** (0.021)
$Mobile_i \times After_{it}$	0.820*** (0.077)	0.222** (0.076)	0.428*** (0.034)	0.403*** (0.033)	0.590*** (0.064)	0.170** (0.061)	0.350*** (0.032)	0.326*** (0.031)
No. of Obs.	5,180	5,180	5,180	5,180	5,180	5,180	5,180	5,108
R-Squared	0.291	0.273	0.308	0.312	0.241	0.229	0.258	0.260

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User control and time fixed effect are included.

Table 32. Main Estimation Results Using Relative Adoption Time

	Num_Card		Card_Day		Num_Quiz		Quiz_Day	
	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC
T-2	0.043 (0.093)	0.043 (0.093)	0.016 (0.040)	0.016 (0.040)	0.012 (0.077)	0.018 (0.077)	0.001 (0.038)	0.002 (0.038)
T+0	1.656*** (0.094)	0.696*** (0.097)	0.804*** (0.038)	0.796*** (0.037)	1.160*** (0.081)	0.484*** (0.079)	0.625*** (0.038)	0.614*** (0.038)
T+1	0.634*** (0.103)	0.134 (0.100)	0.345*** (0.046)	0.300*** (0.044)	0.444*** (0.085)	0.082 (0.081)	0.286*** (0.0433)	0.246*** (0.041)
T+2	0.233* (0.106)	-0.101 (0.100)	0.159*** (0.047)	0.135** (0.046)	0.184* (0.088)	-0.031 (0.085)	0.141** (0.044)	0.122** (0.043)
No. of Obs.	5,180	5,180	5,180	5,180	5,180	5,180	5,180	5,180
R-Squared	0.327	0.285	0.347	0.354	0.269	0.238	0.285	0.290

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User control and time fixed effect are included.

Table 33. Estimation Results Using Fixed Effect

	Num_Card		Card_Day		Num_Quiz		Quiz_Day	
	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC
$After_{it}$	1.192*** (0.084)	0.323*** (0.085)	0.570*** (0.035)	0.549*** (0.034)	0.866*** (0.070)	0.262*** (0.068)	0.456*** (0.034)	0.436*** (0.033)
$Mobile_i \times After_{it}$	0.825*** (0.076)	0.225** (0.075)	0.430*** (0.033)	0.405*** (0.032)	0.595*** (0.063)	0.173** (0.061)	0.352*** (0.031)	0.328*** (0.031)
No. of Obs.	5,180	5,180	5,180	5,180	5,180	5,180	5,180	5,108
Number of user_id_fe	1,036	1,036	1,036	1,036	1,036	1,036	1,036	1,036

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User fixed and time fixed effect are included.

Table 34. Estimation Results Using LA-PSM

	Num_Card		Card_Day		Num_Quiz		Quiz_Day	
	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC	PC+Mobile	PC
$Mobile_i$	-0.088* (0.043)	-0.093* (0.041)	-0.029 (0.019)	-0.030 (0.019)	-0.062+ (0.035)	-0.073* (0.034)	-0.024 (0.018)	-0.022 (0.018)
$After_{it}$	-0.411*** (0.046)	-0.420*** (0.046)	-0.189*** (0.020)	-0.190*** (0.020)	-0.343*** (0.037)	-0.351*** (0.037)	-0.170*** (0.019)	-0.170*** (0.019)
$Mobile_i \times After_{it}$	1.304*** (0.075)	0.712*** (0.074)	0.628*** (0.033)	0.594*** (0.032)	0.966*** (0.063)	0.548*** (0.060)	0.509*** (0.032)	0.481*** (0.031)
No. of Obs.	5,170	5,170	5,170	5,170	5,170	5,170	5,170	5,170
R-Squared	0.342	0.296	0.359	0.361	0.282	0.252	0.292	0.294

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User control and time fixed effect are included.

Table 35. Falsification Test

	Num_Card	Card_Day	Num_Quiz	Quiz_Day
$Mobile_i$	0.043 (0.074)	0.006 (0.032)	0.006 (0.060)	-0.015 (0.029)
$After_{it}$	-0.009 (0.068)	-0.014 (0.030)	-0.007 (0.055)	-0.017 (0.028)
$Mobile_i \times After_{it}$	-0.042 (0.093)	-0.014 (0.040)	-0.005 (0.077)	0.0016 (0.038)
No. of Obs.	2,072	2,072	2,072	2,072
R-Squared	0.583	0.570	0.531	0.536

Notes: (1) Clustered standard errors in parentheses: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1; (2) *Num_Card* and *Num_Quiz* are log transformed; (3) User control and time fixed effect are included.

APPENDIX C

A LIST OF SUTDIED SUBREDDITS

Table 36. List of Studied Subreddits

Subreddit	Adoption Date	Subreddit	Adoption Date	Subreddit	Adoption Date
AdviceAnimals	2013/9/26	Smite	2014/2/4	leagueoflegends	2013/5/24
Art	2013/7/12	TheLastAirbender	2014/10/9	lifehacks	2014/5/1
AskReddit	2013/2/13	TheLeftovers	2014/7/6	lincoln	2013/4/2
Autos	2013/10/30	TheoryOfReddit	2013/6/25	longboarding	2013/8/9
BitMarket	2013/12/1	TumblrInAction	2014/3/16	loseit	2014/3/25
BostonJobs	2013/5/15	TwoXChromosomes	2014/5/10	malefashionadvice	2014/6/24
BostonSocialClub	2013/5/18	UsenetInvites	2013/6/17	memes	2013/9/7
CFB	2013/6/13	Watches	2013/3/18	motorcycles	2014/1/2
Charity	2013/7/4	WildStar	2013/11/17	musicgifstation	2013/4/14
China	2013/7/8	Wordpress	2014/5/4	netflix	2013/10/10
ClashOfClans	2013/10/2	acturnips	2014/4/2	nintendo	2013/7/4
ContagiousLaughter	2014/2/17	airsoftmarket	2014/2/10	nocontext	2013/2/20
DIY	2013/5/9	androidthemes	2013/11/5	nostalgia	2014/1/23
Damnthatsinteresting	2014/4/18	apple	2013/8/18	oddlysatisfying	2014/4/21
Design	2013/10/14	archeage	2014/8/4	offbeat	2013/6/25
DestinyTheGame	2014/9/23	arresteddevelopment	2013/5/12	personalfinance	2014/6/11
Documentaries	2014/5/6	askscience	2013/8/7	pics	2014/5/27
DoesAnybodyElse	2013/2/1	asoiaf	2013/8/26	pokemon	2013/12/20
Entrepreneur	2013/8/20	atheism	2013/7/1	r4r	2013/2/11
Fitness	2013/2/13	beer	2014/2/3	rage	2013/4/29
Forex	2013/8/29	boardgames	2013/7/18	redditgetsdrawn	2013/8/7
Futurology	2014/4/13	books	2013/11/29	runescape	2013/12/6
GetMotivated	2014/7/6	boston	2013/5/17	scifi	2013/8/19
GifSound	2013/4/9	breakingbad	2013/8/12	skyrim	2013/5/28
GiftofGames	2014/6/11	buildapc	2014/3/27	smashbros	2014/6/10
GlobalOffensive	2013/12/1	cats	2013/2/14	snackexchange	2013/11/7
GrandTheftAutoV	2013/8/15	christmas	2013/11/13	snapchat	2013/12/7
GunsAreCool	2013/9/10	circlejerk	2014/3/10	soccer	2013/12/22
IAmA	2013/5/7	computers	2014/1/20	space	2013/7/23
Ijustwatched	2013/4/8	confession	2014/1/24	sports	2014/1/2
IndieGaming	2014/9/3	conspiracy	2013/5/24	startups	2014/3/16
JusticePorn	2013/6/22	cosplay	2013/5/28	summonerschool	2014/4/9

Kikpals	2014/1/29	creepyPMs	2013/5/2	supremeclothing	2014/6/24
LeagueOfGiving	2013/9/20	daddit	2013/2/13	kickstarter	2013/4/17
		dataisbeautiful	2013/6/20	xbox360	2013/5/11
LifeProTips	2013/11/11	dating_advice	2013/2/13	switcharoo	2013/8/6
Loans	2013/10/15	dayz	2013/12/23	sysadmin	2013/5/29
MaddenUltimateTeam	2014/10/20	elderscrollsonline	2013/9/19	technology	2014/2/19
MusicVideos	2014/2/16	explainlikeimfive	2013/9/22	teenagers	2013/11/7
OldSchoolCool	2014/11/20	femalefashionadvice	2013/6/2	television	2013/7/17
Overwatch	2014/11/21	food	2014/5/24	thatHappened	2013/7/6
PS3	2013/4/28	foxes	2013/3/4	thewalkingdead	2013/6/1
PS4	2013/5/24	fullmoviesonyoutube	2013/1/31	tifu	2014/4/28
Pets	2013/9/9	funny	2014/7/2	treemusic	2013/11/28
Poetry	2013/11/9	gadgets	2013/6/24	unitedkingdom	2013/7/2
Pokemongiveaway	2013/11/25	gamedev	2013/8/31	videos	2014/11/29
PropagandaPosters	2013/4/12	gameofthrones	2013/9/20	vita	2013/4/14
RandomKindness	2014/2/13	giftcardexchange	2014/6/24	web_design	2013/10/1
Rateme	2014/11/11	hiphopheads	2014/6/27	windowsphone	2013/11/2
Sherlock	2014/1/8	history	2013/2/16	woahdude	2013/2/13
ShouldIbuythisgame	2013/6/10	hockey	2013/4/29	worldnews	2013/10/13
SkincareAddiction	2013/12/14	iphone	2014/6/2	wow	2013/5/28
jailbreak	2013/5/30				

APPENDIX D

RELATIVE TIME AND SUR MODEL RESULTS

Table 37. The Impact of AutoModerator on Human Moderators' Participation using Relative Time Model

	Moderator Role			User Role
Variables	Num_Mod_Policing	Num_Mod_Explanation	Num_Mod_Suggestion	Num_Mod_Casual_Talk
Relative_Month(t-6)	0.00568 (0.108)	0.0249 (0.0984)	0.106 (0.109)	0.133 (0.102)
Relative_Month(t-5)	-0.0962 (0.103)	-0.100 (0.0961)	0.0411 (0.104)	0.0812 (0.0991)
Relative_Month(t-4)	0.0466 (0.112)	0.0421 (0.103)	0.0827 (0.108)	0.0916 (0.0813)
Relative_Month(t-3)	0.00942 (0.111)	-0.0400 (0.104)	0.121 (0.0958)	0.0581 (0.0831)
Relative_Month(t-2)	-0.0348 (0.0926)	-0.0515 (0.0826)	0.0355 (0.0793)	0.0151 (0.0739)
Relative_Month(t0)	0.247** (0.115)	0.241** (0.109)	0.279*** (0.103)	0.213** (0.0899)
Relative_Month(t+1)	0.202* (0.116)	0.135 (0.110)	0.191* (0.105)	0.187** (0.0823)
Relative_Month(t+2)	0.157 (0.137)	0.0717 (0.133)	0.132 (0.113)	0.154* (0.0843)
Relative_Month(t+3)	0.153 (0.125)	0.120 (0.122)	0.0918 (0.108)	0.0408 (0.0717)
Relative_Month(t+4)	0.229* (0.118)	0.142 (0.120)	0.159 (0.106)	0.0558 (0.0874)
Relative_Month(t+5)	0.223* (0.130)	0.149 (0.127)	0.0709 (0.110)	-0.0392 (0.0865)
Relative_Month(t+6)	0.240* (0.127)	0.185 (0.130)	0.127 (0.110)	-0.0550 (0.0908)
$Num\_User\_Participation_{it}$	0.240*** (0.0678)	0.257*** (0.0688)	0.240*** (0.0755)	0.236*** (0.0668)
Num_Mods	1.464*** (0.140)	1.575*** (0.121)	1.642*** (0.104)	1.796*** (0.0915)
Constant	-2.369*** (0.602)	-2.268*** (0.614)	-1.955*** (0.678)	-1.046* (0.623)
Subreddit fixed effect	Yes	Yes	Yes	Yes

Month fixed effect	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
Number of subreddits	95	95	95	95
N	1,235	1,235	1,235	1,235
R-squared	0.417	0.472	0.512	0.595

Note: (1) Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; (2) For all dependent variable and the number of commenters, we use log transformation (i.e., log(x+1)) to accommodate skewed distribution and zeros.

Table 38. The Impact of AutoModerator on Human Moderators' Participation using SUR Model

	Moderator Role			User Role
Variables	Num_Mod_Policing	Num_Mod_Explanation	Num_Mod_Suggestion	Num_Mod_Casual_Talk
After	0.155*** (0.050)	0.126*** (0.048)	0.099** (0.046)	0.130*** (0.043)
Num_User_Cmts	0.101*** (0.017)	0.111*** (0.016)	0.106*** (0.015)	0.187*** (0.014)
Num_Mods	1.485*** (0.037)	1.636*** (0.036)	1.766*** (0.034)	1.934*** (0.032)
Subreddit fixed effect	Yes	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
Number of subreddits	156	156	156	156
N	3,744	3,744	3,744	3,744
Chi2	15,812.58***	19,325.08***	23,051.35***	35,485.55***

Note: (1) Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; (2) For all dependent variable and the number of commenters, we use log transformation (i.e., log(x+1)) to accommodate skewed distribution and zeros.

APPENDIX E

HYTEROGENEITY IN USERS' PRE-TREATMENT DECLARATION

Table 39. The Moderating Effect of User Pre-treatment Declaration (Attention Allocation)

Variables	(1) Focal_Sub	(2) Other_Politics_Subs	(3) Non_Politics_Subs
After	-0.0257*** (0.00316)	0.0142** (0.00687)	0.0115* (0.00638)
After x With_Flair_Before	0.0302*** (0.0102)	-0.0200 (0.0134)	-0.0102 (0.0114)
Num_Posts_Week	0.0241*** (0.00282)	-0.0147*** (0.00464)	-0.00940** (0.00446)
Constant	-0.0279** (0.0112)	0.554*** (0.0180)	0.474*** (0.0173)
N	23,734	23,734	23,734
User fixed effect	Yes	Yes	Yes
R-squared	0.009	0.001	0.000
Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1			

Table 40. The Moderating Effect of User Pre-treatment Declaration (Discourse Participation)

Variables	(1) Num_Discourses	(2) Num_Discourses_by_Left	(3) Num_Discourses_by_Right	(3) Num_Discourses_by_Undeclared
Afterlis	-1.145*** (0.0472)	-0.719*** (0.0695)	-0.731*** (0.0685)	-1.668*** (0.0793)
With_Flair_Before	0.992*** (0.0654)	0.915*** (0.128)	0.849*** (0.110)	2.499*** (0.195)
After x With_Flair_Before	0.782*** (0.0498)	0.695*** (0.0734)	0.664*** (0.0723)	0.782*** (0.0798)
Num_Posts_Week	0.372*** (0.0332)	0.345*** (0.0465)	0.510*** (0.0471)	0.346*** (0.0482)
Constant	-2.032*** (0.129)	-1.864*** (0.193)	-2.704*** (0.190)	-1.836*** (0.195)
N	39,963	26,298	26,143	26,166
User fixed effect	Yes	Yes	Yes	Yes
Wald Chi	1001.11***	236.99***	328.30***	854.02***
Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1				