Does Identity Disclosure Help or Hurt User Content Generation? Social Presence, Inhibition, and Displacement Effects

Jingchuan Pu

Smeal College of Business
Pennsylvania State University, University Park, Pennsylvania 16802
E-mail: pjchuan@gmail.com

Yuan Chen

School of Information Management and Engineering
Shanghai University of Finance and Economics, Shanghai 200433, China
E-mail: chen.yuan@mail.shufe.edu.cn

Liangfei Qiu

Warrington College of Business
University of Florida, Gainesville, Florida 32611
E-mail: liangfeiqiu@ufl.edu

Hsing Kenneth Cheng

Warrington College of Business
University of Florida, Gainesville, Florida 32611
E-mail: hkcheng@ufl.edu

Abstract

Many user-generated content (UGC) websites are experimenting with disclosing users' identities to increase accountability for the generated content. However, the effects of identity disclosure on users' content-generation behaviors are not well examined. In this study, we address this critical issue by using a natural experiment — a large corporate online community chose to disclose users' identities in one section (the focal section) but not the other (the neighbor section). Our results show that in the focal section, disclosing identity increases social presence and inhibits users' willingness to generate content, resulting in greater effort spent per content but smaller content volume. Surprisingly, we find that users significantly change their content-generation behaviors in the neighbor section, where users remain anonymous. Specifically, identity disclosure has a strong displacement effect: the low-effort content, which is deterred by identity disclosure in the focal section, will be reallocated to the anonymous neighbor section. Furthermore, taking both sections together, we find the content volume increases and content effort exerted on each content decreases overall. These findings demonstrate that identity disclosure is a double-edged sword with regard to user content generation. On the one hand, disclosure motivates users' effort on each content in the focal section. On the other hand, the displacement effect meant this benefit comes at the cost of reducing users' effort per content in the neighbor section.

Keywords*: User-generated content, identity disclosure, social presence, inhibition, displacement effect*

1. Introduction

Anonymity used to be a key feature of user-generated content (UGC) websites, contributing to their popularity but also raising questions about the credibility of the content (Scott and Orlikowski 2014). As the demand for transparency and accountability for UGC is increasing, many websites gradually remove the anonymity feature and begin disclosing users' identities next to their content to boost the credibility and reliability of UGC and to make users more responsible. For instance, reviewers on Amazon can choose to show their identity (Forman et al. 2008); Weibo, the Chinese version of Twitter, has started to verify users' profile and publicly indicate whether each specific user has a verified identity (Wang et al. 2017); Yelp and TripAdvisor have started to integrate the users' accounts with their corresponding Facebook accounts (Huang et al. 2017), and news websites, such as Huffington Post, decided to disclose users' identity next to their comments by linking the Facebook accounts (Fredheim et al. 2015). This integration has moved users from generating content anonymously to doing so under their real identity.

While it appears evident that disclosing users' identities increases the credibility of UGC (Forman et al. 2008), our understanding of how identity disclosure affects users' content-generation behaviors is limited. The success of UGC websites depends entirely on users' voluntary contributions to generate sufficient content and to attract new members (Goes et al. 2016). Regardless of their specific functionalities, those websites consistently share the same concern about how to motivate their users to generate more content and exert greater effort on each content. Image motivation is one of the most important motivations for people to engage in contributing behavior, and it refers to an individual's tendency to be motivated by others' perceptions (Ariely et al. 2009). In the UGC context, the perception that content generation enhances their image significantly motivates users (e.g., Wasko and Faraj 2005, Jabr et al. 2014, Goes et al. 2016, Burtch et al. 2017, Qiu and Kumar 2017).

An important property of image motivation is its dependency on visibility because image is essentially a consequence of what others think (Qiu and Kumar 2017). In this light, how image motivation affects user content generation crucially depends on the level of social presence, which is the degree to which people can sense the others' existence (Short et al. 1976). Identity disclosure can fundamentally increase social presence among users and better allow users to experience others as being psychologically present (Sia et al. 2002), thereby altering how image motivation works. In other words, identity disclosure can affect the image motivation by increasing users' perception of the visibility of their generated content. Given the trend of identity disclosure in the UGC websites and its important role on image motivation, this study aims to examine how identity disclosure influences user content generation in terms of two outcomes: (1) average content length: how identity disclosure increases social presence and affects users' effort exerted on each content, and (2) content volume: how identity disclosure increases social presence and affects the number of generated content.

Existing literature on user content generation mainly focuses on the direct effects of one policy on users' activities where the policy is implemented (e.g., Goes et al. 2016, Chen et al. 2018). However, UGC websites, such as Stack Exchange, often host multiple sections that have different policies for users' activities in different sections. Moreover, users are found to allocate their effort among different activities in the UGC context (Ghose and Han 2011, Singh et al. 2014, Huang et al. 2015). When users generate content in different sections, the policy change in one section can inevitably affect users' behaviors in the other as users can accordingly reallocate activities across sections to better earn image. However, the indirect effects of a policy in one section on users' behaviors in the other section are not well explored. To fill this research gap, we examine how identity disclosure in one section affects user content generation in the other section where users remain anonymous.

Displacement effect describes the phenomenon when one policy is inconsistently implemented in two areas (e.g., the policy is implemented in one area but not the other), the rational individuals would reallocate the policy-related behaviors between these two areas. Displacement effect is well recognized in economics and criminology literature as a policy's indirect effect (Cornish and Clarke 1987, Sun 2005, Lacetera et al. 2012, 2014). When identity disclosure in one section increases social presence at that section, users are likely to reallocate their content-generation activities among different sections to get better image overall through the displacement effect. This study complements the growing literature on user content generation by analyzing both the direct effect and displacement effect of identity disclosure.

This study also provides significant managerial implications. For example, if we find that identity disclosure is detrimental to user content generation, our study will bring attention to an important issue and initiate a discussion on how to balance the benefits from increased content credibility against the losses from content generation. Moreover, if our results demonstrate that after users' identities are disclosed in one section, they also significantly change their behaviors in other section, then we can understand identity disclosure's effects more comprehensively and provide the practitioners with useful guidelines on the other policies. Motivated by the academic and practical implications, we investigate the impacts of identity disclosure when users' identities are disclosed in one section but not in another. Specifically, we seek to answer the following research questions:

How does identity disclosure in one section of a UGC website affect users' content volume and average content length in that section?

How does identity disclosure in one section affect users' content volume and average content length in another section still allowing anonymity?

To answer these research questions, we obtain a unique dataset from a large online community of an international conglomerate subsidiary in China. The community is used in the corporate marketing department. Users are the employees in the marketing department. The community website hosts two sections: a review section that allows users to review the shared cases, and a question and answer (Q&A) section that allows them to ask and answer questions. Users voluntarily generate content by posting reviews and answering questions. For each piece of the generated content, users can be awarded one virtual point. A separate webpage, which is accessible to all users, lists the total points that each user earned this year. The company wanted to understand how identity disclosure benefits the community and decided to experiment with disclosing users' identities next to the content they generated. At the beginning of 2014, the community randomly chose to disclose users' identities in the review section, but not in the Q&A section. This exogenous policy change provides a natural experiment to examine the effects of identity disclosure on user content generation. In the rest of the paper, "the review section" will be referred to as "the focal section", and "the Q&A section" will be referred to as "the neighbor section" interchangeably.

We collect the data on users' content-generation activities both before and after the implementation of identity disclosure. We construct a panel data about users' behaviors at the user-month level and adopt an interrupted time series strategy. Our estimations provide clear comparisons of users' activities during the pre-disclosure period and during the post-disclosure period. The inclusion of user-level fixed effects helps us control for the omitted variables that are either time invariant or that change over time but stay constant across users. Since the decision about which section users' identities would be disclosed is exogenous, our estimations are not vulnerable to selection issues. However, our identification strategy suffers from some alternative explanations. For example, changes in user content generation could have resulted from shifts in case or question topics. To address this concern, we first use the platform's pre-defined question categories and find that the results are consistent across different categories. We then utilize Latent Dirichlet Allocation (LDA), a topic modeling method, to demonstrate that our results are consistent with the controls on the case and question topics. The analyses at the case and question levels further confirm our results. Another possible confounding factor is that users might have changed their behaviors over time even without the identity disclosure. To address this concern, we first construct the treatment and control user group with the new employees and only compare their activities during their participating year because the participating time of the new employees is exogenous to the community's policy. Moreover, we utilize a Regression Discontinuity in Time (RDiT) framework to control for the potential time-varying confounders at the user level (Hausman and Rapson 2018).

Our results indicate that because of the increased social presence in the focal section, which is brought by identity disclosure, users expend greater effort on each content but generate fewer pieces of content in the focal section, as compared to the time before identity disclosure. However, in sharp contrast, we find that users produce more content but expend less effort per content in the neighbor section, implying a significant displacement effect. Furthermore, we observe that after identity disclosure in one section, users generate more content but with less effort per content, taking both sections together. Our results show that the displacement effect is more salient for the users who care more about the image from content volume, and the inhibition as well as displacement effects are less salient for the users who can better earn image from greater effort on each content. Using separate estimations of the content that resulted in virtual points and the content that didn't, we show that the displacement effect is partially caused by the virtual points. Through analyses of different user groups, we also find that the displacement effect is even influential on users who are not directly affected by the identity disclosure.

Our study contributes to the literature in several ways. First, it contributes to the growing literature on the antecedents of UGC (e.g., Wasko and Faraj 2005, Jabr et al. 2014, Qiu and Kumar 2017), exploring identity disclosure as one important antecedent. User content generation is highly motivated by the image motivation, which fundamentally depends on the visibility of users' content. Our study reveals that identity disclosure can enhance social presence and increase users' perceived visibility of their content. Increased social presence motivates users to pay more attention to the image received from each content and exert greater effort per content. Moreover, increased social presence causes an inhibition effect and leads to a smaller content volume. For practitioners, the findings imply that UGC communities should be careful of whether to disclose users' identities, since the removal of the anonymity feature doesn't simply improve the credibility of content but also inevitably changes users' content-generation behaviors via increasing social presence.

Second, our study investigates users' effort reallocation behaviors when they face different policies across different sections, and how users change their behaviors across sections to earn better image. Identity disclosure in the focal section can help motivate higher effort per content in the focal section while reducing users' effort per content in the neighbor section. Moreover, the content volume may be inhibited in one section but encouraged in the other. These compelling findings add new perspectives to the literature on online users' content-generation behaviors. These results also remind practitioners that users' behaviors may differ in different sections of a website. Only a comprehensive understanding of users' activities across different sections can guide the design of efficient platforms to better facilitate UGC.

Finally, this study extends the understanding of the displacement effect from economics and criminology literature to the context of UGC (Sun 2005, Lacetera et al. 2012, 2014, Gonzalez-Navarro 2013). We provide evidence that the displacement effect exists in the UGC websites when the users' identities are disclosed in one section but not in the other. This study sheds new light on the displacement effect by documenting the displacement effect in a virtual rather than geographical space.

The findings of our study also have important practical implications for the UGC websites. While the identity disclosure in the focal section can motivate users exert greater effort on each content by increasing social presence, users tend to decrease content volume in that section because of inhibition effect. Given that UGC websites mainly rely on users' voluntary contribution (Goes et al. 2016), managers of these websites should better comprehensively evaluate the effects of relevant policies. Moreover, our results show that identity disclosure in the focal section can lead to less effort exerted per content in the anonymous neighbor section because of the displacement effect. When there is inconsistent policy implemented in different sections of a UGC website, practitioners should take the possible displacement effect into account and design a better reward system. To sum up, our study demonstrates that identity disclosure can motivate users to exert greater effort per content in the focal section, but this benefit may come at the cost of the decreased content volume in the focal section and the lower users' effort per content in the neighbor section.

The rest of the paper proceeds as follows. In Section 2, we review the related literature. Section 3 formally proposes our hypotheses. Sections 4 and 5 describe our research context and empirical strategy. In Section 6, we discuss our main results, conduct a group of robustness tests, and provide the mechanisms of the observed effects. Finally, Section 7 concludes the paper.

2. Literature Review

In this section, we review three relevant streams of literature: antecedents of UGC, identity disclosure and UGC, and users' effort allocation at UGC context. We also highlight our contributions by comparing our work with past studies.

2.1 Antecedents of UGC

Our study is closely related to the literature about the antecedents of UGC. In online communities, various mechanisms are employed to motivate users to generate more content and exert greater effort on each content (e.g., Wasko and Faraj 2005, Toubia and Stephen 2013, Jabr et al. 2014). Image motivation is one of the most important motivations for people to engage in contributing behavior, and it refers to an individual's tendency to be motivated by others' perceptions (Ariely et al. 2009). Prior studies have documented that image motivations such as perceived status or reputation among peers (Wasko and Faraj 2005, Shen et al. 2015, Goes et al. 2016) and social connections (Zhang and Zhu 2011, Goes et al. 2014, Qiu and Kumar 2017) can effectively motivate user content generation.

For the content generators, a better image in the UGC website can be gained from higher content volume (i.e., *volume-based* image) or greater average effort on each content (i.e., *effort-based* image) (e.g., Wasko and Faraj 2005, Jabr et al. 2014, Goes et al. 2016, Burtch et al. 2017, Qiu and Kumar 2017, Chen et al. 2018).[1] To facilitate users better perceive the volume-based image, online communities use different policies, such as virtual points (Jabr et al. 2014), status system (Goes et al. 2016), or user badges (Chen et al. 2018). Users increase their content volume because it is closely associated with the volume-based image expressed through more virtual points, higher status, or better badges. The pursuit of effort-based image motivates users to exert greater effort on each content. Wasko and Faraj (2005) show that users' better perception about the effort-based image is essential to motivate users to provide high-quality content. Jabr et al. (2014) indicate that when an online community has a mechanism to reflect users' effort on each content, users would produce content with higher quality. Shen et al. (2015) show that reviewers strive to compete for the readers' attention to gain image from each content. The image-related motivation is particularly critical in corporate online community, because the community participation may have direct consequences for users' careers: content-generation activities are closely related to career prospects (e.g., Roberts et al. 2006, Hwang et al. 2015). Hann et al. (2013) find that users' earned image from an open source software community (through a ranking system) is associated with an 18% increase in wages.

Existing studies show that image motivation crucially depends on the perceived visibility, which significantly changes users' content-generation activities. Toubia and Stephen (2013) find that an increase in followers motivates users' pursuit of image and enhances users' content volume in Twitter. Zhang and

[1] Please note that the *effort-based* image here is specified as the image gained from the average effort exerted on each content. It is different from studies in which content volume is the main dependent variable (e.g., Goes et al. 2016).

Zhu (2011) show that greater audience size motivates users to generate more content. Qiu and Kumar (2017) indicate that a large number of followers motivates users to pursue the effort-based image and exert more effort in each content. Users' content-generation activities can be altered when users' perceived visibility increases (Goes et al. 2014). Huang et al. (2017) find that social network integration substantially increases users' social presence and further affects the content volume and emotional expression in each content.

In our research context, users get volume-based image from virtual points. When users are anonymous for their generated content, the image from content generation is likely to be only associated with the content volume and the effort-based image is relatively limited. Identity disclosure in the focal section can affect how users perceive image from their content-generation activities by effectively improving the social presence and the perceived visibility. Our study contributes to this stream of literature in two aspects. We first explore how identity disclosure, which substantially increases social presence, serves as an important antecedent of UGC and affects users' content volume and users' effort per content (measured by average content length). Given the fact that most UGC websites host multiple sections, we then take a step further to investigate how identity disclosure in the focal section can affect users' behaviors in the neighbor section, in which users remain anonymous.

2.2 Identity Disclosure and UGC

Anonymity refers to a state in which identifying information for an acting party is unknown (Hoffman et al. 1999). In practice, identity disclosure, a process to remove the anonymity feature, is gradually adopted by many UGC websites and researchers are starting to study its effects on user content generation. Although the websites under consideration are heterogeneous in terms of functions and identity disclosure strategy, a consistent finding in the literature is that identity disclosure affects user content generation. Leshed (2009) finds that removing anonymity greatly reduces the content volume. Kilner et al. (2005) show that removing anonymity feature leads to fewer antisocial comments and fewer comments in total. Huang et al. (2017) find that social network integration (reduce of anonymity) increases the review volume, and shapes how reviewers use emotional words.

Building on prior research, this study extends our understanding on the effects of identity disclosure on UGC. From the content consumers' perspective, identity disclosure increases the credibility of the content because the attributes of an information source have powerful effects on how people respond to messages (Chaiken 1980, Forman et al. 2008). Identity disclosure can also substantially affect users' content-generation behaviors by increasing social presence and perceived visibility of their content. When their identity is disclosed, users are more likely to publicize socially desirable information (Huberman et al. 2005), generate fewer critical comments (Jessup et al. 1990), and use emotional and uncensored words less frequently (Cho and Kim 2012, Huang et al. 2017). Instead of focusing on the expression or linguistic features, we analyze the question of how identity disclosure affects the social presence and influences user content generation in terms of content volume and the effort expended on each content. Moreover, we focus on users' behaviors in multiple sections where the identity disclosure is only implemented in one of them. Table 1 presents examples of identity disclosure's effects in the literature and their interested variables.

Table 1. Literature on Identity Disclosure and User Content Generation

Literature (research context)	Identity Disclosure Strategy	Interested Aspects				
		Content Volume	Linguistic Feature	Effort per Content	Multiple Sections	Displacement Effect
Kilner and Hoadley 2005 (PlatoonLeader)	Platform's real name policy	√		√		
Leshed 2009 (The Young and Fresh)	Platform's real name policy	√				
Cho and Kim 2012 (Popular website in Korea)	Government's real name policy	√	√			
Omernick and Sood 2013 (TechCrunch.com)	Social network integration		√	√		
Fredheim et al. 2015 (Huffington Post)	Social network integration	√	√	√		
Paskuda and Lewkowicz 2017 (Quora)	Platform's real name policy		√			
Huang et al. 2017 (Yelp and TripAdvisor)	Social network integration	√	√			
This Study (corporate online community)	Platform's real name policy	√		√	√	√

2.3 Users' Effort Allocation in the UGC Context

Users are found to allocate their effort among different activities in the UGC context. For example, Singh et al. (2014) find that readers of a large corporate blog website assign their time on reading the work- and leisure-related posts. Users' work-related blogging activity is also related to their leisure-related content generation (Huang et al. 2015). Moreover, users tend to allocate their effort on different content-generation activities for better image (Kuang et al. 2019). To the best of our knowledge, no prior work has investigated users' activity reallocation across different sections caused by the inconsistent policy change in the UGC context. Our study indicates that because users strive to obtain image from different sections, the identity disclosure in the focal section not only affects users' content-generation activities in that section, but also significantly influences their activities in the neighbor section through the displacement effect.

To accurately assess the effects of crime prevention or economic promotion policies, literature has carefully considered the displacement effect. For example, applying the monetary incentive to volunteers significantly motivates more blood donations in one area, but that incentive also causes smaller number of donations in the neighbor areas or future donations of the same area (Lacetera et al. 2012, 2014). Since consumers tend to displace the future purchase to the promotion period or switch the purchase to the discounted items, post-promotion dip and brand switch are frequently observed in the analyses on marketing promotion effect (Sun 2005). In the criminology literature, the focus of the displacement effect is on individuals' misbehaviors. Displacement effect occurs because the policy increases the risk and the effort of misbehavior in one location but not the other (Cornish and Clarke 1987). For instance, equipping vehicles with Lojack reduces 48% in theft risk at one state, but 18% of the reduction in thefts is displaced to the neighbor states, where the Lojack is not utilized (Gonzalez-Navarro 2013). As the individuals who are influenced by the policy are making decisions at different locations, one important question is always asked: whether the observed effect of the policy at one location is at the cost of its adverse displacement effect at the other locations?

The existing literature mainly focuses on the spatial or temporal displacement effect (Sun 2005, Png et al. 2008, Lacetera et al. 2012, 2014). We extend the understanding of displacement effect to the

UGC context, in which users' activities are separated into different sections of one UGC website. While it is prevalent that same group of users generate content in different sections under different policies, the associated displacement effect is not well analyzed in the literature. If disclosing users' identities in one section significantly affects users' activities in that section, a natural question will be whether this effect is associated with its reverse displacement effect in the other section.

3. Hypothesis Development

Our analyses of identity disclosure focus on users' content-generation activities in both the focal and neighbor sections. In the focal section, we consider how identity disclosure increases social presence, thereby changing users' effort per content and content volume. We then evaluate whether and how identity disclosure in the focal section affects users' content-generation activities in the neighbor section through displacement effect. Figure 1 illustrates our research model.

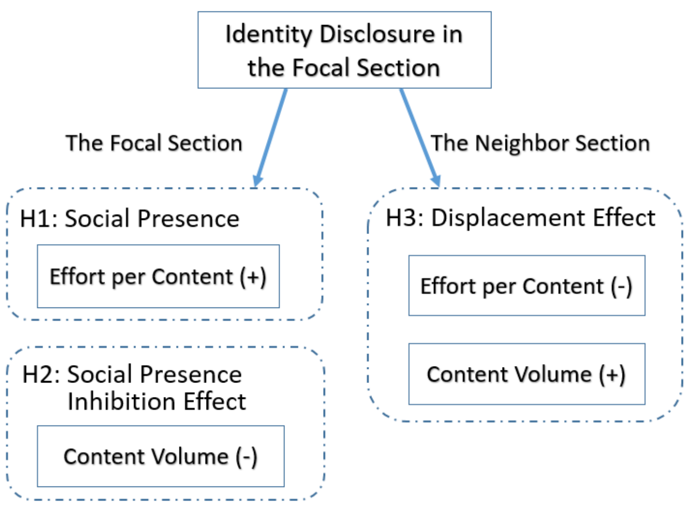


Figure 1. Research Model

The first metric of interest is the content length in the focal section. Image motivation is one of the most important motivations. Individuals derive utility from being judged positively (Gneezy et al. 2012) and they are motivated by others' perceptions (Soetevent 2005, Ariely et al. 2009). How image motivation affects one's behaviors crucially depends on the perceived visibility of his or her behaviors because image is essentially a consequence of what others think (Qiu and Kumar 2017). *Social presence theory* argues that “the degree of salience of the other person in an interaction” can shape individual’s behaviors (Short et al. 1976). Social presence has been found to be an important predictor of user behavior in the computer-mediated communication (Sia et al. 2002, Cobb 2009, Farzan et al. 2011, Cui et al. 2012). Moreover, higher social presence can make individuals value image more, and existing studies show that the level of social presence is positively associated with users’ effort. For instance, Miranda and Saunders (2003) find that higher social presence can facilitate information processing. Studies in social loafing (e.g., Harkins and Petty 1982) suggest that the identified group members exert greater physical or mental effort than those working anonymously.

In the UGC context, users strive to achieve a more positive image by producing more content (volume-based image) or exerting greater effort on each content (effort-based image). One key prerequisite for effort-based image to motivate users is that users can well perceive image from the effort exerted on each content (Wasko and Faraj 2005). Jabr et al. (2014) show that when an online community has a mechanism to reflect users’ effort on each content, users would produce content with higher quality. Wasko and Faraj (2005) show that users’ better perception about the effort-based image is essential to motivate users to provide high-quality content. The reduced anonymity can significantly increase social presence among users, meaning that users will better sense the existence of other individuals (Huang et al. 2017). When a website reduces the anonymity feature, users are likely to adapt their behaviors to their perception of social presence and the visibility of their behaviors (Acquisti and Gross 2006, Jones and Linardi 2014). Since the image motivation fundamentally depends on users’ perceived visibility of their generated content (Qiu and Kumar 2017), we propose that identity disclosure can shape the image motivation by increasing social presence.

When users are anonymous to their content, the perceived effort-based image is limited because there is no effective way for readers to identify the generator of one specific content. In other words, the visibility of users’ effort on each content is limited because of the anonymity feature. Once the users’ identities are disclosed next to the content in the focal section, they would expend greater effort per content as they can realize the better visibility of their content via the increased social presence. The underlying mechanism is similar to that of the social connections' effects on user content generation: once users perceive that the visibility of their contribution increases, they are more inclined to exert greater effort on each contribution to gain a better image (Qiu and Kumar 2017). In the corporate online community, when the generated content can be better observed by other employees in the company (the peers or participants with a higher job rank), users should care more about the image earned through each content because a better image can potentially benefit their career within the company (Hwang et al. 2015).

Moreover, with the increased social presence, users can better sense the existence of other content generators. Shen et al. (2015) show that reviewers strategically change their review activities to compete with other content generators for better attention. To better earn image from the generated content in the focal section, users need to compete with the peers who also generate content in that section and exerting greater effort is an effective way to earn attention. In summary, identity disclosure in the focal section increases social presence in that section, providing users with a good opportunity to better earn effort-based image and compete with peers by exerting greater effort on each content. Accordingly, we propose our first hypothesis:

Hypothesis 1 (H1). *After the identity is disclosed in the focal section, users exert greater effort per content in that section.*

We also expect that identity disclosure can affect the content volume in the focal section. The increased social presence in the focal section, brought by the identity disclosure, may affect users' content volume in different mechanisms.

First, we propose that the identity disclosure can decrease users' content volume through *inhibition effect*. That is, once their identity is disclosed, the social presence increases, and users can better sense the existence of others, leading to the concern about how others perceive their activities. Kling et al. (1999, p.82) mention that on the Web, "people say or write things under the cloak of anonymity that they might not otherwise say or write." Removal of anonymity has been shown to lead to an inhibition effect in different scenarios. Bapna et al. (2016) show that users are less likely to visit others' profiles if their visits are recorded. Andreoni and Bernheim (2009) indicate that as the anonymity decreases, users are likely to conduct the behaviors that might hurt their image. The studies in the group decision support systems (GDSS) consistently report that showing participants' identity restricts the production of comments (e.g., Jessup et al. 1990). The inhibition effect brought by the increased social presence can play an important role in the UGC context. Scott and Orlikowski (2014) point out that anonymity leads users to feel more comfortable and secure, resulting in more contribution in the UGC website. As identity disclosure increases social presence and the perceived visibility of their content, users are likely to worry more about how the others value their content. In this light, apart from increasing the effort spent on each content, users tend to decrease certain types of content. With the loss of anonymity, users are found to express less emotional words (Omernick and Sood 2013, Huang et al. 2017), and show less true and uncensored opinions (Cho and Kim 2012). Fredheim et al. (2015) find that after disclosing commenters' names next to their comments at Huffington Post, the number of comments for the political articles decreases. Kilner and Hoadley (2005) show that identity disclosure in the online community decreases the total number of comments. Leshed (2009) indicates that after showing users' identity next to their comment, the number of comments significantly decreases.

Second, identity disclosure might motivate the users to generate more pieces of content due to the increased social presence. As users better sense the existence of others and the visibility of their content increases, they are more likely to believe that their content can be potentially helpful to others and the higher content volume can bring them better image (Huang et al. 2017). Moreover, identity disclosure may increase the motivation of volume-based image, as users' contribution level can be better perceived through the content volume with the reduction in anonymity. Zhang and Zhu (2011) and Goes et al. (2014) show that the increased content visibility, which brought by a greater audience seize, motivates users to generate more content. Huang et al. (2017) indicate that social network integration (reduction in anonymity) of the review websites increases reviewers' content volume.

In summary, identity disclosure increases social presence, leading countervailing effects on users' content volume. On the one hand, content volume can decrease because users' certain types of content are substantially inhibited with the identity disclosure. On the other hand, content volume may increase because users can more easily earn image from generated content through the increased social presence. We propose that in our research context, inhibition effect will dominate, and users generate fewer pieces of content after identity disclosure.

The community offers the users one virtual point for every generated content, and there is a separate webpage listing users' total virtual points. When the identity is not disclosed next to users' content, this mechanism inevitably makes users overlook the effort on each piece of content, because they can only receive image from the content volume (i.e., volume-based image). Therefore, users are more likely to produce content without spending enough effort on each content (hereafter referred to as "low-effort content"). Once users' identities are disclosed next to their content, which is the similar policy documented in Leshed (2009) and Fredheim et al. (2015), users are likely to reduce some specific types of content (such as low-effort content) that are detrimental to their image. In our context, users in the corporate community can be affected by this inhibition effect because the generated content potentially affect their reputation in the company and job prospect (Huang and Zhang 2016). Meanwhile, the motivation of volume-based image is not particularly amplified in the focal section because the webpage listing users' virtual points offers a good channel for users to earn such image even before identity disclosure. Overall, users will reduce the emotional, socially disagreed, and low-effort content, and our second hypothesis emphasizes this relationship between identity disclosure and content volume:

Hypothesis 2 (H2). *After the identity is disclosed in the focal section, users generate fewer pieces of content in that section.*

In addition to user content generation in the focal section, we are also interested in the effects of identity disclosure on the neighbor section still allowing anonymity. Users tend to allocate their effort among different activities on the UGC websites. Ghose and Han (2011) find a negative relationship between user content generation and consumption behavior on mobile devices. Huang et al. (2015) investigate how users allocate their effort in an enterprise blogging website when facing tradeoffs between work- and leisure-related content. As users need to allocate their effort among different content-generation activities, they are likely to reallocate their activities when the policy of the UGC website changes, and a policy in one section might affect users' behaviors in the other section where such policy is not implemented. In specific, when one policy is inconsistently implemented among different sections in one UGC website, users are likely to be affected by the displacement effect and reallocate their activities among different sections because the reputation risk of generating certain types of content substantially increases in one section but not in the other (e.g., Cornish and Clarke 1987, Lacetera et al. 2012, 2014).

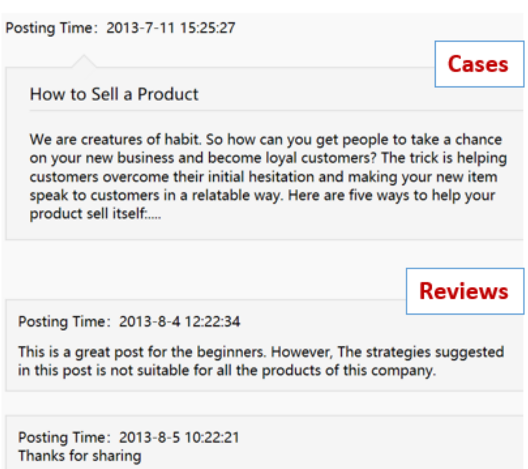
In our context, identity disclosure in the focal section enhances social presence in that section, increasing effort on each content and inhibiting content volume in that section. Users can correspondingly reallocate their activities in both sections to better gain both effort-based and volume-based image overall. Displacement effect can be induced by users' pursuit of image. Users in our setting are the employees within one company and their image in the community can be associated with their explicit payments (Roberts et al. 2006) or the expectation of better employability (Huang and Zhang 2016). The image from total content volume may help them build a better reputation in the company, eventually benefiting their career (Hann et al. 2013). Due to identity disclosure's inhibition effect on the focal section, users' image from content volume (virtual points) is restricted in that section. Although some users may choose not to increase content volume in the neighbor section because they are not mainly motivated by the volume-based image, users who value the volume-based image would choose to produce more in the neighbor section to maintain or gain higher such image (Zhang and Zhu 2011, Goes et al. 2016). Therefore, we should expect that after identity disclosure in the neighbor section, users generate more content in the neighbor section on average.

Although content generators are likely to increase content volume in the neighbor section, the effort-based image is not critical in that section because they remain anonymous to their content. Also, users can displace the low-effort content, which is inhibited by the identity disclosure in the focal section, to the neighbor section. Therefore, users are inclined to decrease the effort per content in the neighbor section. Our third hypothesis reflects this users' reallocation of activities caused by the displacement effect.

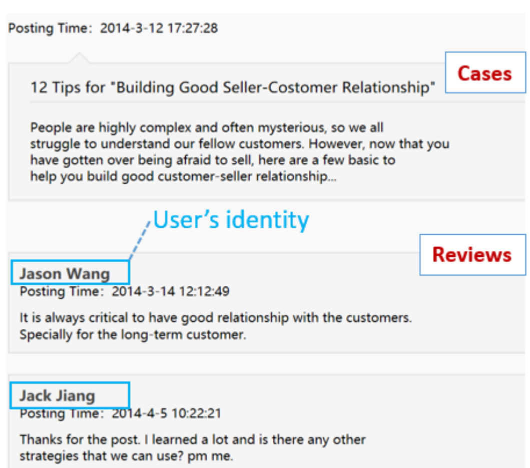
Hypothesis 3 (H3). *After the identity is disclosed in the focal section, users generate more content but exert less effort per content in the neighbor section.*

4. Research Context and Data

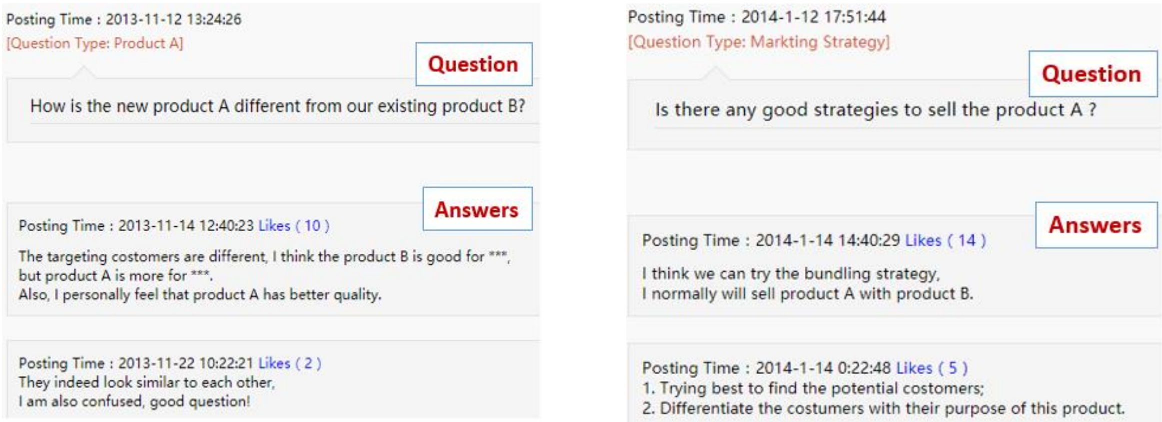
The community in our study hosts two sections: a review section and a Q&A section. The community is used in a corporate marketing department, and the participants are the employees in that department. The company in our study mainly sells consumer goods in China and the employees are distributed in different provinces across the country. In the review section, users provide opinions about the shared cases, which are mainly about the company's products and marketing advices. In the Q&A section, users ask and answer the questions about the company's products or specific marketing strategies. Users are awarded one virtual point for each piece of their generated content. A separate webpage, which is accessible to all users, lists the total points that each user earned this year. At the beginning of 2014, the company randomly chose to disclose users' identities in the review section, but not in the Q&A section. We focus on the users' activities from January 2013 to December 2014. Panels A and B in Figure 2 illustrate the layout of review section in 2013 and 2014, respectively. From the interviews with the manager of this community, we learn that the only systematic change happened within our study period is that the users' identities are disclosed in the review section since the beginning of 2014 (Figure 2, Panels A and B). The layout of Q&A section remains consistent (Figure 2, Panels C and D). This exogenous policy change provides us with a unique natural experiment to detect the effects of identity disclosure on user content generation.



Panel A. Review Section in 2013



Panel B. Review Section in 2014



Panel C. Q&A Section in 2013 Panel D. Q&A Section in 2014

Figure 2. Translated Screenshots of the Online Community

To understand how identity disclosure affects user content generation, we focus on the active users who joined the community prior to the study period and continued to post content throughout the study period. Consequently, the user sample of our main analyses includes 591 users. In addition, our dataset consists of 36,107 reviews and 429,857 answers. For each content, we collect information on the posting time and its full textual content. We use the number of characters to quantify the content length, because in Chinese, characters form the basic unit of meaning.

5. Estimation Strategy

We are interested in how users change their behaviors in both answering the questions and reviewing the cases when their identities are only disclosed in the review section. We assess the effects of identity disclosure in terms of two outcomes: (1) content volume: the total volume of content, and (2) effort on each content: the average length of generated content. Both outcomes are important to the communities' success. We focus on users' answer and review behaviors because they are users' major content-generation activities. On average, our sampled users contribute 32.851 answers or reviews per month, but only 0.802 cases or questions per month during our study period. In Online Appendix A, we discuss the effects of identity disclosure on users' case and question generation.

Difference-in-differences (DID) at the user-section level seems to be useful because we can utilize the users' activities in the focal section as treatment and their behaviors in the neighbor section as control.

However, with the existence of policy externalities, this DID estimation can produce biased estimates of policy impact (Miguel and Kremer 2004). The basic challenge is that whatever treatment in users' activities in the treated section (the focal section) also affects their activities in the control section (the neighbor section), and these are no longer valid counterfactual observations. Essentially, DID estimation precludes actual estimation of externalities unless there is a set of observations subject to externalities, and a set of observations that is not so that they can play the role of counterfactual. However, because of the potential displacement effects, such counterfactual observations do not exist. For these reasons, we do not use DID estimation. Instead, we use an interrupted time series strategy in which the counterfactual is given by the observations occurring before the policy intervention (Gonzalez-Navarro 2013, Zhang and Zhu 2011). We construct a panel data set such that each unit is a user and each time period is one calendar month.

Dependent variables. The main outcomes we are interested in are users' content volume and the average effort that users expend per content. Apart from assessing the effects of identity disclosure on user content generation in different sections, we also evaluate its overall impacts by taking two sections together. We measure users' content volume with total number of content user i generates in each month t: (1) the number of reviews (i.e., $NbrReview_{it}$), (2) the number of answers (i.e., $NbrAnswer_{it}$), and (3) the number of reviews plus the number of answers (i.e., $NbrContent_{it}$). We use the average length of the content generated by the user i in each month t to quantify his or her effort per content: (1) the average length of reviews (i.e., $LenReview_{it}$), (2) the average length of answers (i.e., $LenAnswer_{it}$), and (3) the average length of reviews and answers (i.e., $LenContent_{it}$). We estimate the effects of identity disclosure on users' content generation with the following models:

$$\log(Y_{it}+1) = \beta_0 + \beta_1 Year2014_t + Control_{it} + \sum_{m=2}^{12} \gamma_m MonthDummy_m + \delta_i + \epsilon_{it} \qquad (1)$$

where i and t indicate a user and a month respectively. Y_{it} is the aforementioned dependent variables of user i at month t. Because our dependent variables are all positive values, we take natural logarithms of all dependent variables plus one before including them in the estimation.

The variable $Year2014_t$ is our main covariate indicating whether users' identities are disclosed in the focal section, and it equals 1 if the observation is for users' activities in 2014, and 0 otherwise. We also control for a series of observed covariates, $Control_{it}$. We first include the variable $UserExp_{it}$, which is the total number of reviews and answers user i generated up until month t, to quantify the content generation experience that user i has in the community. Moreover, when the dependent variable is content volume (i.e., the number of reviews or the number of answers), we include the number of cases or questions generated in the community in month t. For instance, when the dependent variable is $NbrAnswer_{it}$, we include the variable $QuesCount_t$, which is the total number of questions generated by all users in month t. We include calendar month dummies to control for unobserved time trend and potential seasonal effects. We incorporate the user-level fixed effects δ_i to control for unobservable time-invariant user-level characteristics. We cluster the error terms at the user level to account for autocorrelation over time (Sun and Zhu 2013). Last, β_0 is an intercept and ϵ_{it} is a mean-zero random error term.

Table 2.1 describes the descriptions and summary statistics of our main variables. 591 sampled individuals are used for our main analysis, and the balanced panel data set contains 14,184 user-month pairs. The average pieces of content contributed by users is 32.851 per month, in which the users generate 2.546 reviews and 30.306 answers. Moreover, to reduce the skewness of our control variable $UserExp_{it}$, we take the natural logarithms of it before estimating our model.[2] Table 2.2 shows the average content length before and after identity disclosure in the review section, and it also presents the content length comparisons with t-tests at the content level. The results provide some model-free evidences.

[2] We obtain consistent results when $UserExp_{it}$ is included in the estimation without the log transformations.

Table 2.1 Main Variables at User-Month Level

Variable	Description	Mean	Std. Dev.	Min	Max	N
$Year2014_t$	A dummy variable which indicates whether the observation describes the time period in 2014: 1 means month *t* is in 2014, 0 otherwise.	0.5	0.5	0	1	14,184
$NbrReview_{it}$	Number of reviews user *i* posts at month *t*.	2.546	10.862	0	316	14,184
$NbrAnswer_{it}$	Number of answers user *i* posts at month *t*.	30.306	52.609	0	244	14,184
$NbrContent_{it}$	The sum of number of answers and reviews user *i* posts at month *t*.	32.851	55.096	0	470	14,184
$LenReview_{it}$	The average length of reviews user *i* posts at month *t*.	31.518	14.369	7	255	4,017
$LenAnswer_{it}$	The average length of answers user *i* posts at month *t*.	36.152	22.376	3	255	6,814
$LenContent_{it}$	The average length of answers and reviews user *i* posts at month *t*.	34.428	19.300	3	255	7,664
$UserExp_{it}$	The total number of answers and reviews user *i* posts by month *t*.	513.125	704.934	0	4,993	14,184
$CaseCount_t$	Number of new cases in the community at month *t*.	133.208	142.534	6	662	14,184
$QuesCount_t$	Number of new questions in the community at month *t*.	669.417	624.507	22	1,899	14,184
$PostCount_t$	Number of new questions and cases in the community at month *t*.	802.625	534.381	252	1,938	14,184

Note: The number of observations for $LenReview_{it}$, $LenAnswer_{it}$ and $LenContent_{it}$ is less than the number of observations for the other variables. Because our observation is on user-month level, for the month that a user did not generate any content, the count variables equals 0, but the variables of content length are missing.

Table 2.2 Statistics at Content Level

	Before January 2014	After January 2014	Difference
Average Review Length	27.485	30.525	3.040***
Average Answer Length	37.246	29.600	-7.646***
Average Length of Reviews and Answers	34.036	29.628	-4.408***

Note. The unit of T-tests is at the content level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

6. Empirical Results

6.1 Main Results

We present the estimation results of our main analyses in Table 3. When the dependent variable is average length of reviews, β_1 is positive and statistically significant (Table 3, Model 4), supporting the social presence effect of identity disclosure: the average length of content increases after users' identities are disclosed in the focal section (H1). However, when the dependent variable is the number of reviews, β_1 is significantly negative (Table 3, Model 1), supporting the hypothesis that users generate a smaller number of reviews as their identity is disclosed (H2). That is, identity disclosure increases user's average effort but inhibits the content volume. Meanwhile, the coefficient of $Year2014_t$ is significantly negative when the dependent variable is $LenAnswer_{it}$ and positive when the dependent variable is $NbrAnswer_{it}$ (Table 3, Models 2 and 5), indicating that identity disclosure has a strong displacement effect: in the neighbor section still allowing anonymity, users generate more content but exert less effort on each content (H3).

Table 3. The Effects of Identity Disclosure on User Content Generation

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent var.	$NbrReview_{it}$	$NbrAnswer_{it}$	$NbrContent_{it}$	$LenReview_{it}$	$LenAnswer_{it}$	$LenContent_{it}$
$Year2014_t$ (β_1)	-0.266*** (0.057)	0.604*** (0.088)	0.807*** (0.078)	0.104*** (0.021)	-0.156*** (0.025)	-0.053*** (0.019)
$UserExp_{it}$	-0.024*** (0.015)	0.371*** (0.045)	0.351*** (0.043)	-0.030** (0.014)	-0.101*** (0.013)	-0.079*** (0.011)
$CaseCount_t$	-0.001*** (0.000)	-	-	-	-	-
$QuesCount_t$	-	0.001*** (0.000)	-	-	-	-
$PostCount_t$	-	-	0.000*** (0.000)	-	-	-
Month Dummies	YES	YES	YES	YES	YES	YES
User Fixed	YES	YES	YES	YES	YES	YES
R-squared	0.280	0.557	0.544	0.532	0.558	0.510
Observations	14,184	14,184	14,184	4,017	6,814	7,664

Note. Robust standard errors clustered by each user are in parentheses. The unit of analyses is at user-month level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Our results show that users expend greater effort per content but generate a smaller content volume in the focal section after identity disclosure, while reverse changes are observed in the neighbor section. Since identity disclosure has opposing effects in two sections, it is intriguing to investigate its net effect. After taking both sections together, we find that the average length of generated content significantly decreases, and the content volume increases overall after identity disclosure in the focal section (Table 3, Models 3 and 6). The results indicate that in the UGC context, as identity disclosure jointly affects users' content-generation activities on both the focal and neighbor sections, users tend to generate greater number of content but less effort on each content overall.

6.2 Robustness Checks

Our main analyses indicate that after their identity is disclosed in the review section, users contribute fewer reviews but exert greater effort per review, and they generate more answers but exert less effort per answer in the Q&A section. We further run several additional tests to ensure that our main results are robust. Table 4 summarizes these tests.

Table 4. Summary of Robustness Checks

Findings	Robustness Checks	Results
Our main results are robust after controlling for the potential reciprocity effect.	Additional control variables	Table 5
Effects of identity disclosure do not simply come from the variation of question or case topics.	(1) Analyses on each question category (2) Additional control variables generated from LDA	Tables 6 and 7
Our main results are robust with different measurements of users' effort per content.	(1) Different measurements of users' effort per content at user-month level (2) Analyses at case or question level	Tables 8 and 9
Our main results are robust with different empirical specifications.	(1) Analyses at case or question level (2) Negative binomial panel model	Tables 9 and 10
Effects of identity disclosure do not simply come from the time trend around the policy implementation time.	(1) Analyses on the new employees (2) RDiT	Tables 11 and 12

6.2.1 Reciprocity Effect

In this study, we are particularly interested in the effects of identity disclosure on users' review and answer activities, and we analyze these effects by using a natural event and comparing users' reviews and answers before and after that event. However, user content generation is also potentially affected by the reciprocity effect (Johnson et al. 2014). For example, some users may need help and ask questions in the Q&A section, and they provide more answers as a return. In this sub-section, we generate several additional variables to control for the potential reciprocity effect. For example, when we analyze users' answer activities, we re-estimate our models with one additional control variable, $QuesCount_{it}$, which stands for the log-transformed number of questions generated by user i in month t. This variable quantifies the intensity of reciprocity effect in the Q&A section because the number of questions posted by user i represents how much help needed from the community and the extent that he or she needs to reciprocate with answers. In this light, we also generate $CaseCount_{it}$ and $PostCount_{it}$, which represent the number of cases and the number of cases plus the number of questions generated by user i in month t, respectively.

We report the results in Table 5. For all 6 dependent variables, the estimation of β_1 is quantitatively consistent with our main results in Table 3. The identity disclosure affects users' review and answer activities after we take into account the potential reciprocity effect.

Table 5. The Effects of Identity Disclosure on User Content Generation with Controlling for Reciprocity Effect

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent var.	$NbrReview_{it}$	$NbrAnswer_{it}$	$NbrContent_{it}$	$LenReview_{it}$	$LenAnswer_{it}$	$LenContent_{it}$
$Year2014_t$ (β_1)	-0.212*** (0.054)	0.411*** (0.085)	0.692*** (0.075)	0.097*** (0.022)	-0.150*** (0.025)	-0.052*** (0.019)
$UserExp_{it}$	-0.027*** (0.015)	0.401*** (0.043)	0.373*** (0.041)	-0.029** (0.014)	-0.103*** (0.013)	-0.079*** (0.011)
$CaseCount_t$	-0.001*** (0.000)	-	-	-	-	-
$QuesCount_t$	-	0.001*** (0.000)	-	-	-	-
$PostCount_t$	-	-	0.000*** (0.000)	-	-	-
$CaseCount_{it}$	0.657*** (0.063)	-	-	-0.023* (0.014)	-	-
$QuesCount_{it}$	-	0.930*** (0.054)	-	-	-0.018** (0.008)	-
$PostCount_{it}$	-	-	0.904*** (0.046)	-	-	-0.006 (0.008)
Month Dummies	YES	YES	YES	YES	YES	YES
User Fixed	YES	YES	YES	YES	YES	YES
R-squared	0.280	0.596	0.586	0.532	0.558	0.510
Observations	14,184	14,184	14,184	4,017	6,814	7,664

Note. Robust standard errors clustered by each user are in parentheses. The unit of analyses is at user-month level. **p* < 0.1; ***p* < 0.05; ****p* < 0.01.

6.2.2 Question and Case Topics

After the identity is disclosed in the review section, users significantly change their content-generation behaviors for both reviews and answers, even though they remain anonymous in the Q&A section during our study period. One alternative explanation is that the changes in users' behaviors purely stem from the shift of the question or case topics, because the production of answers and reviews is closely related to the available questions and cases in the community. In the following analyses, we first utilize a simple category-based classification to differentiate the questions, and then use LDA to discover the latent topics of the questions and cases and control for them in our regression.

Table 6. The Effects of Identity Disclosure on Users' Answer Activities in Different Categories

Category	Marketing Strategy		Product 1		Product 2	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent var.	$NbrAnswer_{it}$	$LenAnswer_{it}$	$NbrAnswer_{it}$	$LenAnswer_{it}$	$NbrAnswer_{it}$	$LenAnswer_{it}$
$Year2014_t$ (β_1)	0.889*** (0.091)	-0.198*** (0.027)	0.161*** (0.058)	-0.116*** (0.028)	0.145*** (0.033)	-0.085*** (0.034)
$UserExp_{it}$	0.360*** (0.043)	-0.089*** (0.013)	0.219*** (0.028)	-0.087*** (0.016)	0.073*** (0.015)	-0.077*** (0.017)
$QuesCount_t$	0.001*** (0.000)	-	0.001*** (0.000)	-	0.000*** (0.000)	-
$MarketingQ Ratios_t$	-0.590*** (0.140)	-	-	-	-	-
$Product1Q Ratios_t$	-	-	2.900*** (0.205)	-	-	-
$Product2Q Ratios_t$	-	-	-	-	0.783*** (0.126)	-
Month Dummies	YES	YES	YES	YES	YES	YES
User Fixed	YES	YES	YES	YES	YES	YES
R-squared	0.570	0.571	0.519	0.514	0.417	0.428
Observations	14,184	6,334	14,184	5,275	14,184	4,421

Note. Robust standard errors clustered by each user are in parentheses. The unit of analyses is at user-month level. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

We first use the categories that are pre-defined by the community. The questions are classified into nine categories such as "marketing strategy," "leisure," and the categories about the company's products. If the users' answer activities change simply because of the shift of question interests, we should observe that the answer patterns change inconsistently across different question categories. To check this alternative explanation, we construct the dependent variables as the number of and the average length of the answers for each category and test them with equation (1). We include one additional control variable as the portion of the questions from that particular category to the total number of the questions. We only report the results on "marketing strategy," "product 1," and "product 2" because they are the biggest categories in terms of the number of questions. Results are represented in Table 6: across all categories, the coefficient of $Year2014_t$ is consistently positive and significant when the dependent variable is the number of answers

(Table 6, Models 1, 3, and 5) and it is significantly negative when the dependent variable is the average answer length (Table 6, Models 2, 4, and 6). Users change their answer activities consistently across different question categories and thus this systematical change is not due to the shift of question interests.

It is also conceivable that the pre-defined categories do not quantify other attributes of questions and cases, which might be associated with the users' answer and review behaviors. For example, users may expend greater effort on answers because the questions become lengthier or the question topics become more suitable for lengthy answers. To rule out such explanation, we use text mining and let the data speak for themselves. Following Bapna et al. (2018), we use LDA to discover latent topics of the available questions and cases, and further control for them in our regression. The LDA is a Bayesian statistical and information retrieval technique. The input to the LDA is a set of documents. In this study, the input is the questions or cases produced in 2013 and 2014. For example, if we need the latent topics of the questions, the input of LDA is the text of the questions and the output is a predefined number (K) of topics from those questions and a posterior topic distribution over all K topics for each question. For each question j, the LDA model provides a K-vector, $< T_{j,1}, T_{j,2}, ..., T_{j,K} >$, where $T_{j,k}$ represents the weight of question j associated with topic k. Also, for each question j and topic k, $0 < T_{j,k} < 1$ and $\sum_{k=1}^{K} T_{j,k} = 1$. The larger the weight of a given topic, the more likely it is that the question j is associated with that topic.

We first run the LDA model with the text of questions. The number of topics is the only degree of freedom in fitting an LDA model. We test the models with K = 5, 10, and 15 topics, and find that 5 is the optimal number of latent topics in terms of the posterior log-marginal likelihood (Griffiths and Steyvers 2004, Nam et al. 2017, Bapna et al. 2018). We observe that the five topics reflect either the products sold by the company or the marketing strategies used by the employees.

After getting the topic distribution of the questions, we confirm our results by taking into account the question topics. In our main regression model, each observation corresponds to a user-month pair. Now we can add the variables describing the topics of questions. For example, if J questions are generated in month t, the trend of the topics of questions in month t can be described by the aggregated topics of those questions < $\overline{T_1}, \overline{T_2}, \ldots, \overline{T_K}$ >, where $\overline{T_k} = \frac{1}{J}\sum_{j=1}^{J} T_{j,k}$. We include the average length of the questions, as well as the trend of topics 1, 2, 3 and 4 as control variables, and re-estimate our model. Topic 5 is our reference group. Including all five topics in our model would lead to perfect collinearity since $\sum_{k=1}^{5} \overline{T_k} = 1$. We follow these steps for the analyses on the cases, and we find that 5 is the optimal number of latent topics. Table 7 illustrates the results, and the estimations of β_1 are qualitatively consistent with our main results.

Table 7. The Effects of Identity Disclosure on Users Content Generation with Controlling for Question and Case Topics

Dependent var.	(1)	(2)	(3)	(4)
	NbrReview$_{it}$	*NbrAnswer$_{it}$*	*LenReview$_{it}$*	*LenAnswer$_{it}$*
Year2014$_t$ (β_1)	-0.389*** (0.080)	0.446*** (0.169)	0.183*** (0.039)	-0.575*** (0.089)
UserExp$_{it}$	-0.017 (0.015)	0.444*** (0.046)	-0.023 (0.014)	-0.087*** (0.013)
CaseCount$_t$	-0.001*** (0.000)	-	-	-
QuesCount$_t$	-	-0.000*** (0.000)	-	-
$\overline{CasesTopic_{1,t}}$	1.094** (0.356)	-	-0.314 (0.274)	-
$\overline{CasesTopic_{2,t}}$	1.218*** (0.192)	-	-0.018 (0.108)	-
$\overline{CasesTopic_{3,t}}$	1.626*** (0.324)	-	1.033*** (0.190)	-
$\overline{CasesTopic_{4,t}}$	0.748*** (0.151)	-	0.005 (0.091)	-
$\overline{QuestionsTopic_{1,t}}$	-	-5.222*** (0.404)	-	-0.329 (0.218)
$\overline{QuestionsTopic_{2,t}}$	-	-4.221*** (0.407)	-	-0.120 (0.264)
$\overline{QuestionsTopic_{3,t}}$	-	-7.550*** (0.573)	-	0.927*** (0.255)
$\overline{QuestionsTopic_{4,t}}$	-	-7.573*** (0.760)	-	-0.002 (0.416)
AvgCaseLen$_t$	0.003 (0.007)	-	0.005 (0.005)	-
AvgQueslen$_t$	-	0.002*** (0.000)	-	0.001*** (0.000)
Month Dummies	YES	YES	YES	YES
User Fixed	YES	YES	YES	YES
R-squared	0.286	0.557	0.538	0.566
Observations	14,184	14,184	4,017	6,814

Note. Robust standard errors clustered by each user are in parentheses. The unit of analyses is at user-month level. *p < 0.1; **p < 0.05; ***p < 0.01.

6.2.3 Alternative Measurements of Effort per Content

In our main analyses, we use the average length of the content generated by user *i* to quantify his or her effort. We find that because of the increased social presence and displacement effect, the identity disclosure causes the users to exert greater effort on each review but less effort on each answer. In this section, we leverage different measurements of users' effort per content, and show that our results are robust.

We first quantify users' effort per content by using the log-transformed average number of non-stop words (i.e., *NonStop*). The non-stop words represent the words that have some concrete meaning, such as noun, verb, and adjective. This variable quantifies the effort on each content by measuring the amount of the concrete information in each content. We test it with equation (1), and the results are reported in Table 8, Models 1 - 3. The coefficient of $Year2014_t$ is significantly positive when the dependent variable is the average number of non-stop words for the reviews (Table 8, Model 1), but is negative and significant when the dependent variable quantifies the average number of non-stop words for the answers (Table 8, Model 2) or answers and reviews overall (Table 8, Model 3). These results indicate that after the identity disclosure in the review section, users tend to provide more concrete information in each review but provide less concrete information in each answer and each content overall.

Secondly, we use two variables, $AvgLike_{it}$ and $BstAnswerRatio_{it}$, to quantify users' average effort on answers. In the Q&A section, content readers click the "like" button if they think the answer benefits the community. $AvgLike_{it}$ is the average number of likes the user *i* received from their answers generated in month *t*. For each question, the question owner can select one answer as the best answer to this question. $BstAnswerRatio_{it}$ is the ratio of number of best answers to the total number of answers generated by the user *i* in month *t*. Those two variables quantify the users' effort spent on the answers through reactions of the readers, and they can quantify how well the answer is matched to the question. We test those dependent variables with equation (1). The coefficients of $Year2014_t$ are negative and significant for $AvgLike_{it}$ (Table 8, Model 4) and $BstAnswerRatio_{it}$ (Table 8, Model 5). That is, after the identity is disclosed in the review section, the answers provided by the users in the Q&A section receive a smaller number of "likes" and have lower chance to be chosen as the best answer.

Thirdly, we assess users' effort exerted on each answer by analyzing whether their answers are related to the corresponding questions. We generate another dependent variable $AnsQueRelevance_{it}$, which represents the average relevance of user i's answers to the corresponding questions in month t. A higher value of this variable can represent greater effort invested in the answers. We calculate the relevance of one answer to its corresponding question by first conducting an LDA model by using all the question and answer texts as input and converting them to K-vectors, where K is the predefined number of latent topics. We then calculate the cosine similarity between the vector of the answer and the vector of its corresponding question. We use the average cosine similarity, $AnsQueRelevance_{it}$, as average effort user i invested in the answers in month t. The coefficients of $Year2014_t$ are negative and significant for $AnsQueRelevance_{it}$ (Table 8, Model 6), indicating that after identity disclosure in the focal section, users generate the answers that are less relevant to their corresponding questions. We only present the result when $K = 5$ because it is the optimal number of latent topics in terms of the posterior log-marginal likelihood. The results are quantitatively consistent when $K = 10$ or 15.

Finally, the users can respond to the questions that already had answers or the questions without any answers. Providing an answer to the question that had no response normally requires more effort than responding to the question that already had some solutions. In this light, we generate a new variable $FirstAnswerRatio_{it}$, which represents the ratio of number of user i's answers for the questions that didn't have any response to his or her total number of answers generated in month t. Our result (Table 8, Model 7) shows that after identity disclosure in the review section, users are less likely to provide the first answer to the questions.

Table 8. The Effects of Identity Disclosure on User Content Generation with Alternative Measurements

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent var.	$NonStop\_Review_{it}$	$NonStop\_Answer_{it}$	$NonStop\_Content_{it}$	$AvgLike_{it}$	$BstAnswerRatio_{it}$	$AnsQueRelevance_{it}$	$FirstAnswerRatio_{it}$
$Year2014_t$ (β_1)	0.091*** (0.021)	-0.160*** (0.023)	-0.062*** (0.019)	-0.027*** (0.011)	-0.006* (0.004)	-0.018*** (0.005)	-0.016** (0.007)
$UserExp_{it}$	-0.027** (0.014)	-0.099*** (0.012)	-0.074*** (0.011)	0.004 (0.006)	0.0001 (0.002)	-0.002 (0.003)	0.002 (0.002)

Table 8. The Effects of Identity Disclosure on User Content Generation with Alternative Measurements (Continued)

Month Dummies	YES	YES	YES	YES	YES	YES	YES
User Fixed	YES	YES	YES	YES	YES	YES	YES
R-squared	0.488	0.562	0.495	0.142	0.166	0.196	0.157
Observations	4,020	6,814	7,665	6,352	6,352	6,813	6,814

Note. Robust standard errors clustered by each user are in parentheses. The unit of analyses is at user-month level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

In our main model, we quantify users' effort on the platform with the average length of reviews or answers generated by a specific user. However, in a given month *t*, a user may review multiple cases and answer different questions. In the previous robustness checks, we incrementally add the attributes of published cases or questions as additional controls to the panel data model. Although such strategy helps us control for the average topics of cases and questions, the aggregation process doesn't precisely represent the exact characteristics of one specific case or question. To overcome this shortcoming, we construct the data at the case level and the question level, respectively. For example, to estimate the effects of identity disclosure on the effort exerted on the answers, we use the following model:

$$\log(\overline{Y_j} + 1) = \alpha_0 + \alpha_1 PostYear2014_j + Control_j + \sum_{m=2}^{12} \gamma_m Month_m + \epsilon_j \quad (2)$$

in which each observation stands for a question. $\overline{Y_j}$ represents the average length of the answers for question j, and $PostYear2014_j$ is a binary variable indicating whether this question is posted in 2014. A set of control variables, $Control_j$, quantify the features of this question: the topics (LDA) and the length of the question j. We also include the calendar month dummy to represent the posting time. Essentially, we use the questions generated in 2013 as control group and the ones generated in 2014 as treatment group. Confounding variables that might affect the dependent variable, the users' effort for question j, and our key independent variable $PostYear2014_j$, could potentially bias our results. Therefore, we further test the results using propensity score matching (PSM) combined with a regression method (OLS).

By using PSM, we construct proper control and treatment groups of matched questions and ensure that these groups are balanced on the observable characteristics of these questions: topic, length, and posting month. We use the nearest neighbor matching algorithm to match each treated question with the most similar questions in the control group (closest propensity score) in terms of the topics and length of the question and the generation month dummy. A logistic regression with these characteristics helps us obtain the predicted propensity score to each question. We then re-estimate equation (2) using the new sample created by the PSM procedure. Following these steps, we conduct the similar analyses for the cases to uncover the effects of identity disclosure on users' reviewing activities. The results are presented in Table 9. The coefficient of $Year2014_j$ is significantly positive when the dependent variable is the average length of reviews to the case (Table 9, Models 1 and 2) and significantly negative when the dependent variable is the average length of answers to the question (Table 9, Models 3 and 4). Therefore, the estimations on the question or case level are consistent with the results of our main analyses.

Table 9. The Effects of Identity Disclosure on Users Content Generation, Case or Question Level

Dependent var.	$LenReview_j$		$LenAnswer_j$	
	(1) OLS	(2) PSM + OLS	(3) OLS	(4) PSM + OLS
$Year2014_j$ (α_1)	0.016** (0.007)	0.018** (0.008)	-0.268*** (0.007)	-0.276*** (0.003)
$CasesTopic_{1,j}$	-0.004 (0.024)	-0.034 (0.034)	-	-
$CasesTopic_{2,j}$	0.064*** (0.011)	0.102*** (0.019)	-	-
$CasesTopic_{3,j}$	0.034*** (0.008)	0.091*** (0.028)	-	-
$CasesTopic_{4,j}$	0.035*** (0.009)	0.011 (0.018)	-	-
$QuestionsTopic_{1,j}$	-	-	-0.048*** (0.005)	-0.061*** (0.005)
$QuestionsTopic_{2,j}$	-	-	-0.028*** (0.005)	-0.029*** (0.005)
$QuestionsTopic_{3,j}$	-	-	0.007 (0.005)	0.022*** (0.005)
$QuestionsTopic_{4,j}$	-	-	0.053*** (0.004)	0.077*** (0.005)
$LenCase_j$	0.055*** (0.006)	0.057*** (0.011)	-	-
$LenQuestion_j$	-	-	0.054*** (0.004)	0.062*** (0.003)
Month Dummies	YES	YES	YES	YES
R-squared	0.149	0.359	0.361	0.368

Note. The unit of analyses is at question or case level. We use nearest propensity score matching with replacement to construct the balanced control and treatment group. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

6.2.4 Different Empirical Specifications

We use the log-transformed values of dependent variables in our main analyses because they are all positive and count or quasi-count in nature. Moreover, we use an interrupted time series strategy in which the treated observations are users' activities after the identity disclose and control observations are users' activities before the time intervention. In this section, we show that our results are consistent with the original scale of the dependent variables, and our main results do not simply come from the time trend of user content generation.

We use the original scale of our interested dependent variables (not log-transformed) and implement a negative binomial panel regression model to account for the over-dispersion (Hausman et al. 1984). Results are reported in Table 10, and they are consistent with those reported in Table 3.

Table 10. The Effects of Identity Disclosure on Users' Content Generation, Binomial Panel Regression Model

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent var.	$NbrReview_{it}$	$NbrAnswer_{it}$	$NbrContent_{it}$	$LenReview_{it}$	$LenAnswer_{it}$	$LenContent_{it}$
$Year2014_t$	-0.603*** (0.072)	0.666*** (0.061)	0.589*** (0.035)	0.067*** (0.015)	-0.231*** (0.013)	-0.128*** (0.012)
$UserExp_{it}$	0.052*** (0.016)	0.357*** (0.012)	0.207*** (0.015)	0.002 (0.009)	-0.055*** (0.007)	-0.031*** (0.006)
$CaseCount_t$	-0.002*** (0.000)	-	-	-	-	-
$QuesCount_t$	-	0.001*** (0.000)	-	-	-	-
$PostCount_t$	-	-	0.000*** (0.000)	-	-	-
Month Dummies	YES	YES	YES	YES	YES	YES
User Fixed	YES	YES	YES	YES	YES	YES
Observations	13,464	13,632	13,824	3,968	6,792	7,653

Note. Standard errors in parentheses. The unit of analyses is at user-month level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Our main analyses track the activities as existing reviewers before and after the identity disclosure to reduce the inconsistence issues in the estimation. However, one concern is that our main results simply come from users' systematical changes over time. That is, the error term ϵ_{it} is correlated with the treatment term $Year2014_t$ of equation (1) in terms of time-varying effects. We rule out this concern with (1) analyses on the new employees in 2013 and 2014, and (2) RDiT framework (Hausman and Rapson 2018).

We first focus on the content-generation behaviors of new employees. It is evident that whether and when an employee joins the company are exogenous to whether identity is disclosed in the online community. We differentiate these new users by which year they participated in the community (i.e., 2013 or 2014). We estimate the effects of the identity disclosure on the new users with the following model.

$$\log(Y_i + 1) = \beta_0 + \beta_1 JoinYear2014_i + \beta_2 Age_i + \beta_3 Gender_i + \epsilon_i \quad (3)$$

where the dependent variables include the log-transferred number of reviews, number of answers, number of both reviews and answers, average length of reviews, average length of answers, and average length of reviews and answers in user *i*'s participating year. For example, if user *i* joined the company in 2013, Y_i represents his or her content-generation activities in 2013. By doing this, we make sure that if user *i* joined the company in 2013, the dependent variables only quantify his or her activities when the identity is not disclosed. $JoinYear2014_i$ is a binary variable which equals 1 if user *i* joins the platform in 2014, the year when the users' identities are disclosed in the review section. We control for user *i*'s demographic information such as gender and age. Table 11 reports the results. Comparing with the users who joined the community in 2013, the users who joined in 2014 generate fewer reviews but exert greater effort on each one (Table 11, Models 1 and 4) and provide more answers but with lower effort on each answer in their first year (Table 11, Models 2 and 5). These results confirm that it is the systematic community policy change that alters the user content generation (identity disclosure is the only change during the study period).

Table 11. The Effects of Identity Disclosure on New Users' Content Generation, New Employees

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent var.	$NbrReviewYear_i$	$NbrAnswerYear_i$	$NbrContentYear_i$	$LenReviewYear_i$	$LenAnswerYear_i$	$LenContentYear_i$
$JoinYear2014_i$	-0.615*** (0.222)	1.400*** (0.287)	1.094*** (0.265)	0.076* (0.045)	-0.214*** (0.072)	-0.087 (0.067)
Age_i	-0.081*** (0.015)	-0.128*** (0.021)	-0.127*** (0.019)	0.013** (0.005)	0.043*** (0.007)	0.041*** (0.006)
$Gender_i$	-0.044 (0.222)	0.324 (0.298)	0.284 (0.275)	-0.070 (0.049)	-0.096 (0.069)	-0.078 (0.065)
Observations	202	202	202	157	188	200

Note. Robust standard errors are in parentheses. The unit of analyses is at user level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

We then use an RDiT to estimate the effects of identity disclosure on user content generation. Specifically, we modify equation (1) and estimate the following regression model:

$$\log(Y_{it}+1) = \beta_0 + \beta_1 Year2014_t + Control_{it} + f(x_t) + \delta_i + \epsilon_{it} \qquad (4)$$

in which x_t is measured as number of months that month t from the January of 2014 (the first month that users' identities are disclosed). In our main analyses, calendar month dummies help us control for unobserved seasonal effects within one year. However, one may question that our observed effects simply come from the time trend around the policy implementation time. RDiT helps us alleviate this concern because the potential time-varying confounders are assumed to change smoothly around the policy implementation time, and the potentially endogenous relationship between ϵ_{it} and time interruption can be controlled by a flexible function $f(x_t)$ (Hausman and Rapson 2018). In our research context, we believe that the relationship between ϵ_{it} and time does not change discontinuously around January of 2014 and a smooth function can describe the non-linear time trend of users' content-generation activities. We follow the existing literature (Davis 2008, Auffhammer and Kellogg 2011) and set $f(x_t)$ as high-order polynomials of x_t. This setting creates a smooth and flexible relationship between ϵ_{it} and x_t, and thus helps us control for the time-varying users' behavior patterns around the policy implementation time. We test the model with $f(x_t)$ as forth-order polynomial, fifth-order polynomial, and sixth-order polynomial of x_t. Table 12 reports the estimations from 18 separate regressions with regression model (4). Each cell provides the coefficients correspond to $Year2014_t$. The results are consistent with our main results.

Table 12. The Effects of Identity Disclosure on Users Content Generation, RDiT

Dependent var.	$NbrReview_{it}$	$NbrAnswer_{it}$	$NbrContent_{it}$	$LenReview_{it}$	$LenAnswer_{it}$	$LenContent_{it}$
Forth-order Polynomial	-0.175*** (0.039)	2.000*** (0.067)	1.85*** (0.063)	0.159*** (0.016)	-0.147*** (0.018)	-0.037** (0.015)
Fifth-order Polynomial	-0.146*** (0.042)	2.097*** (0.066)	1.933*** (0.062)	0.171*** (0.016)	-0.127*** (0.019)	-0.024* (0.015)
Sixth-order Polynomial	-0.116*** (0.042)	2.067*** (0.068)	1.870*** (0.065)	0.173*** (0.016)	-0.114*** (0.018)	-0.016 (0.015)

Note. Robust standard errors clustered by each user are in parentheses. The unit of analyses is at user-month level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

6.3 Mechanisms

Our main results and robustness tests consistently indicate the casual effects of identity disclosure on user content generation: after disclosing users' identity in the focal section, users exert greater effort per content but produce less content in that section, but they tend to generate more pieces of content and exert less effort per content in the neighbor section. In this section, we use a group of additional tests to verify our theorization. Table 13 summarizes these analyses.

Table 13. Summary of Tests for Mechanisms

Findings	Tests for Mechanisms	Results
Displacement effect is associated with the extent of users' interests on the volume-based image.	(1) Moderation analysis (2) Subsample analysis	Tables 14 and 15
Inhibition and displacement effects are associated with the extent of users' interests on the effort-based image.	Moderation analysis	Tables 16 and 17
Effects of increased social presence are associated with users' need to compete for attention (effort-based image).	Moderation analysis	Table B.1
Displacement effect is associated with the existence of virtual point (volume-based image).	Analyses on the content that doesn't provide virtual points	Table B.2
Displacement effect can even influence users who did not generate any reviews.	Moderation analysis	Table B.3

6.3.1 Heterogeneous Effects Analyses

In the hypothesis development, we propose that identity disclosure affects users' content-generation activities in both the focal and neighbor sections because it changes how users perceive the image gained from the content-generation activities in two sections. For a comprehensive understanding of these observed effects, we explore the moderating role of user characteristics regarding how much users care about the effort-based and volume-based images.

We propose that identity disclosure in the focal section can change users' content-generation activities because users start to gain effort-based image in the focal section. It affects their activities in the neighbor section due to the displacement effect: the users gain volume-based image and in response to the identity disclosure in the focal section, they tend to displace the inhibited content to the neighbor section. If the intensity of displacement effect on the neighbor section depends on how users care about volume-based image, we should expect that the displacement effect is more salient for the users who care more about the image from the content volume. We utilize total content, regardless of section, prior to our study period to capture the extent of users' interest in the volume-based image, because the page that lists each user's virtual points is available even prior to our study period.

We change our equation (1) by adding one interaction: $Year2014_t \times NbrContentBy2012_i$, where $NbrContentBy2012_i$ is the log-transformed number of content that user i generated by end of year 2012. Table 14 reports the results. The coefficient of the interaction term is significantly positive when the dependent variable is $NbrAnswer_{it}$ (Table 14, Model 2) and negative when the dependent variable is $LenAnswer_{it}$ (Table 14, Model 5). Moreover, the coefficient of the interaction term is significantly negative when the dependent variable is $NbrReview_{it}$ (Table 14, Model 1), indicating that the identity disclosure's inhibition effect and displacement effect are intensified when the users care more about the volume-based social image. Also, the coefficient of the interaction term is significantly negative when the dependent variable is $LenReview_{it}$ (Table 14, Model 4), implying that the identity disclosure's effect on review length is negatively moderated by the extent of users' interests on the volume-based image. The results complement our main analyses by showing the intensity of identity disclosure's effect is closely related to users' pursuit of volume-based image.

Table 14. DID Estimations on Users with Different Content Volume Prior to the Study Period

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent var.	$NbrReview_{it}$	$NbrAnswer_{it}$	$NbrContent_{it}$	$LenReview_{it}$	$LenAnswer_{it}$	$LenContent_{it}$
$Year2014_t$	0.095 (0.071)	-1.295*** (0.122)	-0.765*** (0.115)	0.298*** (0.066)	0.098 (0.069)	0.188*** (0.058)
$Year2014_t \times NbrContentBy2012_i$	-0.069*** (0.019)	0.384*** (0.027)	0.548*** (0.037)	-0.034*** (0.011)	-0.044*** (0.011)	-0.042*** (0.009)
$UserExp_{it}$	-0.067*** (0.014)	0.608*** (0.038)	0.318*** (0.026)	-0.055*** (0.014)	-0.131*** (0.011)	-0.109*** (0.013)
$CaseCount_t$	-0.001*** (0.000)	-	-	-	-	-

Table 14. DID Estimations on Users with Different Content Volume Prior to the Study Period (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent var.	$NbrReview_{it}$	$NbrAnswer_{it}$	$NbrContent_{it}$	$LenReview_{it}$	$LenAnswer_{it}$	$LenContent_{it}$
$QuesCount_t$	-	0.001*** (0.000)	-	-	-	-
$PostCount_t$	-	-	0.000*** (0.000)	-	-	-
Month Dummies	YES	YES	YES	YES	YES	YES
User Fixed	YES	YES	YES	YES	YES	YES
R-squared	0.284	0.581	0.560	0.535	0.561	0.514
Observations	14,184	14,184	14,184	4,017	6,814	7,664

Note. Robust standard errors clustered by each user are in parentheses. The unit of analyses is at user-month level. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Moreover, we conduct subsample analyses to further understand how displacement effect is caused by the volume-based image. We estimate identity disclosure's effects on the users who generated a small number of content prior to our study period and on the users who generated a large number of content before our study period, respectively. In operationalizing this, we first rank the users based on the number of content they generated before our study period. We treat users in the bottom 50^{th} percentile as those who are not especially interested in the reputation from the content volume: There are a total of 296 users, and on average they generated only 30.990 pieces of content prior to our study period. In contrast, we consider users in the top 50^{th} percentile as those who are concerned about the volume-based image: There are a total of 295 users and on average they produced 471.719 pieces of content prior to our study period.

Table 15 presents the results of the subsample analyses. Panel A shows the results for the users in the bottom 50^{th} percentile regarding the number of content generated prior to our study period, and Panel B indicates the results for the users in the top 50^{th} percentile. The results for the users in the top 50^{th} percentile are quantitatively consistent with our main results in terms of the identity disclosure's effects on users' activities in the both focal and neighbor sections (Table 15, Models 5-8). Interestingly, our results show that the users in the bottom 50^{th} percentile do not significantly change their behaviors in the neighbor section (Table 15, Models 2 and 4), although the sign of the estimation is consistent with our main results. These results further validate our conjecture that users change their behaviors in the neighbor section partially because of the volume-based image.

Table 15. Subsample Analyses

Dependent var.	Panel A (Users in Bottom 50th Percentile)				Panel B (Users in Top 50th Percentile)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Nbr Review$_{it}$	*Nbr Answer$_{it}$*	*Len Review$_{it}$*	*Len Answer$_{it}$*	*Nbr Review$_{it}$*	*Nbr Answer$_{it}$*	*Len Review$_{it}$*	*Len Answer$_{it}$*
Year2014$_t$	-0.149*** (0.049)	0.090 (0.090)	0.117*** (0.045)	-0.018 (0.050)	-0.202** (0.097)	0.721*** (0.133)	0.111*** (0.025)	-0.143*** (0.033)
UserExp$_{it}$	-0.021** (0.014)	0.494*** (0.045)	-0.027 (0.021)	-0.131*** (0.019)	-0.214*** (0.059)	0.655*** (0.101)	-0.048*** (0.023)	-0.157*** (0.027)
CaseCount$_t$	-0.001*** (0.000)	-	-	-	-0.001*** (0.000)	-	-	-
QuesCount$_t$	-	0.001*** (0.000)	-	-	-	0.001*** (0.000)	-	-
Month Dummies	YES	YES	YES	YES	YES	YES	YES	YES
User Fixed	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	0.260	0.561	0.566	0.572	0.257	0.550	0.505	0.553
Observations	7,104	7,104	1,476	2,762	7,080	7,080	2,541	4,052

Note. Robust standard errors clustered by each user are in parentheses. The unit of analyses is at user-month level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

We propose that identity disclosure increases users' effort on each content in the focal section because social presence increases users' self-awareness and the sense about others' existence. If identity disclosure affects users' effort in each content through how users earn effort-based image, we should expect that the users who care more about the reputation from the effort per content are less affected by the inhibition and displacement effect. However, users' average content length prior to the study period does not capture how they value the effort-based image well. Because the online community was anonymous during that period. Instead, we utilize user *i*'s job rank and number of co-workers to represent the extent that he or she can earn better image from greater effort per content.

The company has well-differentiated job ranks. We expect that the users with higher job rank would be less affected by the inhibition and displacement effects because higher job rank stands for higher social status in the company. The users with higher status can derive more reputation from greater effort (Lowenthal and Dennen 2017). Based on the job description, we divide the job ranks into 4 different levels. We use the users with the fourth level job (the highest level) as the benchmark, and we create three binary variables *JobRank_j$_i$* to describe user *i*'s job rank, where *j* ranges from 1 to 3 and a higher *j* represents a higher job rank. We alter equations (1) to (5) by adding three interactions: $Year2014_t \times JobRank\_j_i$. The estimations of the interactions indicate how different is the change of content-generation activities of the users with job rank 1 to 3 from the change of activities of the users with the highest job rank.

$$\log(Y_{it}+1) = \beta_0 + \beta_1 Year2014_t + \sum_{j=1}^{3} \beta_j Year2014_t \times JobRank\_j_i + X_{it} + \sum_{m=2}^{12} \gamma_m MonthDummy_m + \delta_i + \epsilon_{it} \quad (5)$$

Table 16 reports the regression results. The coefficients of $Year2014_t \times JobRank1_i$ and $Year2014_t \times JobRank2_i$ are significantly negative when the dependent variable is $NbrReview_{it}$ (Table 16, Model 1). However, the coefficients of the interactions are significantly positive when the dependent variable is $NbrAnswer_{it}$ (Table 16, Model 2). Comparing to the users with the highest job rank, the users with lower job ranks are more affected by the inhibition and displacement effects.

Table 16. DID Estimations on Users with Different Job Rank

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent var.	$NbrReview_{it}$	$NbrAnswer_{it}$	$NbrContent_{it}$	$LenReview_{it}$	$LenAnswer_{it}$	$LenContent_{it}$
$Year2014_t$	-0.045 (0.098)	-0.027 (0.119)	0.349 (0.263)	0.311 (0.216)	0.016 (0.118)	0.087 (0.084)
$Year2014_t \times JobRank1_i$	-0.295*** (0.097)	0.759*** (0.276)	0.573** (0.271)	-0.227 (0.219)	-0.155 (0.120)	-0.131 (0.082)
$Year2014_t \times JobRank2_i$	-0.252*** (0.096)	0.554** (0.278)	0.375 (0.276)	-0.174 (0.220)	-0.204* (0.107)	-0.140** (0.084)
$Year2014_t \times JobRank3_i$	-0.028 (0.097)	0.686** (0.326)	0.663** (0.322)	-0.197 (0.235)	0.221 (0.181)	0.188 (0.099)
$UserExp_{it}$	-0.022 (0.017)	0.365*** (0.046)	0.348*** (0.045)	-0.035*** (0.013)	-0.102*** (0.013)	-0.080*** (0.011)
$CaseCount_t$	-0.001*** (0.000)	-	-	-	-	-
$QuesCount_t$	-	0.001*** (0.000)	-	-	-	-
$PostCount_t$	-	-	0.000*** (0.000)	-	-	-
Month Dummies	YES	YES	YES	YES	YES	YES
User Fixed	YES	YES	YES	YES	YES	YES
R-squared	0.270	0.554	0.520	0.517	0.550	0.498
Observations	13,128	13,128	13,128	3,837	6,518	7,324

Note. Robust standard errors clustered by each user are in parentheses. The unit of analyses is at user-month level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

The company in our study is composed of different local regions in China and users are from different regions. One local region is far away from the other ones. Thus, the users might be well connected with the co-workers from the same region but not with the ones from a different region. We use the number of co-workers from the same region as a proxy for social connectedness. Here, the number of same-region co-workers can stand for the number of readers with strong social ties (e.g., Goes et al. 2014). Therefore, a higher number of co-workers can help users derive better image from greater effort per content. We change equation (1) by adding interaction $Year2014_t \times CoWorkers_i$, where $CoWorkers_i$ is the log-transformed number of co-workers that user *i* has in his or her local region. Table 17 reports the results. The coefficient of the interaction term is significantly positive when the dependent variable is $NbrReview_{it}$ (Table 17, Model 1) and significantly negative when the dependent variable is $NbrAnswer_{it}$ (Table 17, Model 2).

Table 17. DID Estimations on Users with Different Number of Co-workers

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent var.	$NbrReview_{it}$	$NbrAnswer_{it}$	$NbrContent_{it}$	$LenReview_{it}$	$LenAnswer_{it}$	$LenContent_{it}$
$Year2014_t$	-0.679*** (0.130)	1.076*** (0.226)	1.086*** (0.220)	0.055** (0.045)	-0.199*** (0.055)	-0.053*** (0.038)
$Year2014_t \times CoWorkers_i$	0.102*** (0.026)	-0.118** (0.363)	-0.069 (0.050)	0.013 (0.011)	0.012 (0.013)	0.000 (0.009)
$UserExp_{it}$	-0.017 (0.015)	0.363*** (0.044)	0.347*** (0.043)	-0.030** (0.014)	-0.101*** (0.013)	-0.079*** (0.011)
$CaseCount_t$	-0.001*** (0.000)	-	-	-	-	-
$QuesCount_t$	-	0.001*** (0.000)	-	-	-	-
$PostCount_t$	-	-	0.000*** (0.000)	-	-	-
Month Dummies	YES	YES	YES	YES	YES	YES
User Fixed	YES	YES	YES	YES	YES	YES
R-squared	0.285	0.558	0.544	0.532	0.558	0.510
Observations	14,184	14,184	14,184	4,017	6,814	7,664

Note. Robust standard errors clustered by each user are in parentheses. The unit of analyses is at user-month level. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

With a higher job rank or a higher number of co-workers, users can gain image from greater effort per review more easily after the identity is disclosed. Hence, for the users with higher job ranks and the users who have higher number of co-workers, the number of reviews is less likely to decrease, and the number of answers is less likely to increase after their identities are publicly shown next their reviews. That is, the inhibition effect and the displacement effect are less salient for them. The results help us verify how effort-based image can moderate the effects of identity disclosure. The analyses on different job ranks and number of co-workers also imply that the users with different job-related attributes may care about the volume-based image differently, and users' status in the company is closely associated with their content-generation activities on the corporate online community (Hann et al. 2013).

6.3.2 Additional Analyses

We conduct several additional analyses to further unveil the underlying mechanisms. First, we use the average length of all reviews on the community to quantify users' needs to compete with peers and earn image. Our results indicate that identity disclosure's effect on review length is related to users' needs to compete for the attention with other users in the focal section. Second, we leverage the answer category that doesn't provide any virtual point and show that the displacement effect can be intensified by the existence of virtual point (volume-based image). Finally, we find that users change their answer activities even when they did not generate any reviews before the identity disclosure, indicating that identity disclosure may affect users' behaviors in the neighbor section via peer effects. We report the details of model and regression results in Online Appendix B.

7. Discussion

The current understanding of identity disclosure is limited to its inhibition effect on the UGC's linguistic features (e.g., Cho and Kim 2012) or how disclosing identity affects content readers' reactions (Forman et al. 2008). To the best of our knowledge, this study is the first to investigate the impact of identity disclosure on users' content-generation behaviors in multiple sections in terms of both content volume and effort exerted per content. We utilize a unique dataset from a large corporate online community and take advantage of a natural experiment in which identity disclosure only occurs in the focal section, but not in the neighbor section. We find that after identity disclosure in the focal section, (1) users exert greater effort on each content but produce a smaller number of content in the focal section due to the increased social presence and inhibition effect, (2) users exert lower effort on each content but generate more pieces of content in the neighbor section because of the displacement effect, and (3) taking users' activities in two sections together, after identity disclosure in the focal section, users exert lower effort on each content and generate greater number of content overall. These findings demonstrate that identity disclosure is a double-edged sword. On the one hand, identity disclosure increases users' effort on each content in the focal section. On the other hand, this benefit comes at the cost of reduced users' effort per content in the neighbor section. The content volume is significantly inhibited in the focal section, and users tend to displace the inhibited content to the neighbor section.

Our subsequent analyses use both pre-defined categories and LDA to identify the case and question topics. These approaches provide cleaner identification and suggest that the changes in users' activities do not result purely from the shift of case and question topics. We also analyze the behaviors of new employees and adopt an RDiT framework to demonstrate that the change in users' content-generation behaviors is due to the initiation of identity disclosure. We further check the mechanism by which identity disclosure affects user content generation. Our analyses indicate that the users who care more about the volume-based image are more affected by the displacement effect. We also find that the users who can earn more image from higher effort exerted per content, those with higher job positions or with higher numbers of co-workers, are less affected by the inhibition and displacement effects. Moreover, the displacement effect is partially caused by the virtual points of the community, which represent users' volume-based image. Users are more likely to generate answers that can provide the virtual points (though they seem to answer more questions to some extent regardless of the reward).

7.1 Literature Contribution and Managerial Implications

This study contributes to several important streams of literature. First, it adds to recent discussions on the antecedents of UGC (e.g., Wasko and Faraj 2005, Jabr et al. 2014, Qiu and Kumar 2017). Image motivation is one of the most important motivations in the UGC context, and it crucially depends on the social presence among users and perceived visibility of their content. We propose that identity disclosure serves as an important antecedent of UGC, increasing social presence and affecting the image motivation. Disclosing users' identities next to their content makes users perceive higher visibility of their behaviors on each content and thus motivates greater effort exerted per content. Meanwhile, the increased social presence also leads to an inhibition effect that decreases users' willingness to contribute.

Second, our study broadens the understanding of the pros and cons of identity disclosure with respect to the supply side of UGC. Different from the studies on how identity disclosure increases the credibility of content (Forman et al. 2008), our analyses show that such disclosure will inevitably affect users' content-generation activities through the increased social presence. Specifically, after identity disclosure in the focal section, users tend to generate fewer pieces of content and exert greater effort on each content in that section. Our results complement current understanding of the relationship between identity disclosure and content volume (e.g., Huang et al. 2017) by showing that when users' identities shown next to their content (similar policy in Leshed 2009 and Fredheim et al. 2015) and users can earn image from different activities in different sections, identity disclosure's inhibition effect can dominate and eventually decrease users' content volume. Moreover, identity disclosure can also affect user content generation in the neighbor sections. This study provides insights on the necessity of evaluating the platform policy from the perspectives of both content readers and generators.

Lastly, while the displacement effect has been documented in the literature with regard to the reallocation of restricted behavior to a different geographic place (e.g., Cornish and Clarke 1987, Freeman et al. 1996, Gonzalez-Navarro 2013), our study enriches this stream of literature by showing such an effect in the UGC context. In this study, we focus on the effects of identity disclosure in the focal section on user content generation across multiple sections and the results indicate that, through the displacement effect, identity disclosure in the focal section can also significantly affect users' activities in the neighbor section. Our analyses imply that one policy may not only have the observed effects on the policy-targeting section, but also result in some effect on users' other activities because they can reallocate their activities among different sections to better gain image overall. Moreover, our results indicate that the displacement effect of identity disclosure is partially caused by the current reward mechanism of the platform. The displacement effect might also indirectly affect the users through peer effects.

Our results have important managerial implications for UGC websites and companies relying heavily on UGC. Given the current trend of integrating social network accounts with UGC platform accounts and of UGC websites disclosing users' identities alongside their content, it is important for practitioners to better understand all the possible effects of identity disclosure. For example, we show that identity disclosure results in a significant increase of social presence, which inevitably affects user content generation. While identity disclosure can motivate users to exert more effort per content and inhibit content generation in the focal section, its displacement effect should not be overlooked, either. We also show that the displacement effect is related to the provided virtual points. Thus, practitioners making policies for these platforms might consider ameliorating unintended displacement effect through a better designed reward system. More generally, past work has documented the effects of motivation policies on user content generation (Goes et al. 2016, Chen et al. 2018), but the effects of such policies have only been considered within the section where the policies are implemented. Practitioners should be careful about the overall effects of such policies on users' activities across all sections.

7.2 Limitations and Future Research

This study has several limitations that present future research opportunities. First, when we analyze the displacement effect of identity disclosure, we do not pay specific attention to the potential differences between the motivation of review generation and the motivation of answer generation. For example, the review generation can be motivated for better attention and the answer generation can be motivated by the intention to help others. However, we argue that the difference between review and answer generation would not significantly affect our results. Because image motivation is found as one of the major motivations in both Q&A website (Chen et al. 2018) and review website (Shen et al. 2015, Huang et al. 2017), and how users care about the image from content volume and each content (our main interests) should be similar in their review and answer generation.

Second, our results are limited in their generalizability, given the fact that we focus on a corporate online community. Comparing to users in the websites like TripAdvisor and Yelp, users in our context are highly socially connected and therefore are more likely to be affected by the image motivation. We believe that as long as users care about the effort-based image (Goes et al. 2014, Qiu and Kumar 2017), identity disclosure next to their content can motivate their effort exerted on each content. Moreover, in our context, a separate webpage exists to indicate users' virtual points which effectively represent users' volume-based image. When users' identities are disclosed next to their content (identity disclosure in our study), the motivation of volume-based image is not likely to be particularly amplified in the focal section, and the inhibition effect dominates. This result is different from the result based on the identity disclosure policy in other research context where the motivation of volume-based image can be amplified and lead to a greater content volume (Huang et al. 2017). Therefore, we believe that our results on content volume in the focal section are more suitable for the setting when identity disclosure motivate users to appreciate more effort-based image but not the volume-based image (e.g., Leshed 2009, Kilner and Hoadley 2005, Fredheim et al. 2015). Also, our results show that the existence of virtual point further intensifies the displacement effect. Although the policies based on the content volume, such as virtual point or rank systems, are widely used (Goes et al. 2016, Jabr et al. 2014), we need to interpret the results on displacement effect cautiously in other contexts. Albeit the limitation on the generalizability, the community in this study provides us with a unique natural experiment to investigate the casual impacts of identity disclosure in both the focal and neighbor sections.

Third, we mainly focus on the effects of identity disclosure on user content generation and do not pay much attention on users' content-consumption activities, such as their browsing behaviors. It would be interesting for future research to consider the potential effects of identity disclosure on users' content-consumption activities. Finally, our study focuses mainly on users' content volume and the effort exerted on each content because we are particularly interested in how identity disclosure affects users via the increased social presence. The existing studies have shown that some linguistic features such as the emotional words usage would be inhibited after identity disclosure (Fredheim et al. 2015, Huang et al. 2017), and textual content analyses can be used to investigate the impact of identity disclosure on internal knowledge creation and sharing patterns within the organization in future research.

References

Acquisti A, Gross R (2006) Imagined communities: Awareness, information sharing, and privacy on the Facebook. *International workshop on privacy enhancing technologies Privacy Enhancing Technologies* (Springer, Berlin), 36–58.

Ariely D, Bracha A, Meier S (2009) Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review* 99(1):544-55.

Andreoni J, Bernheim BD (2009) Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica* 77(5):1607–1636.

Auffhammer M, Kellogg R (2011) Clearing the air? The effects of gasoline content regulation on air quality. *American Economic Review* 101(6):2687–2722.

Bapna R, Ramaprasad J, Shmueli G, Umyarov A (2016) One-way mirrors in online dating: A randomized field experiment. *Management Science* 62(11):3100–3122.

Bapna S, Benner MJ, Qiu L (2018) Nurturing online communities: An empirical investigation. *MIS Quarterly* Forthcoming.

Burtch G, Hong Y, Bapna R, Griskevicius V (2017) Stimulating online reviews by combining financial incentives and social norms. *Management Science* 64(5):2065-2082.

Chaiken S (1980) Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology* 39(5):752–766.

Chen W, Wei X, Zhu KX (2018) Engaging voluntary contributions in online communities: A hidden markov model. *MIS Quarterly* 42(1):83-100.

Cho D, Kim S (2012) Empirical analysis of online anonymity and user behaviors: The impact of real name policy. Proc. *45th Hawaii International Conference on System Sciences*, 3041-3050.

Cornish DB, Clarke RV (1987) Understanding crime displacement: An application of rational choice theory. *Criminology* 25(4):933-948.

Cobb SC (2009) Social presence and online learning: A current view from a research perspective. *Journal of Interactive Online Learning* 8(3):241–254.

Cui G, Lockee B, Meng C (2012) Building modern online social presence: A review of social presence theory and its instructional design implications for future trends. *Education and Information Technologies* 18(4):661-685.

Davis LW (2008) The effect of driving restrictions on air quality in Mexico City. *Journal of Political Economy* 116(1):38–81.

Farzan R, Dabbish LA, Kraut RE, Postmes T (2011) Increasing commitment to online communities by designing for social presence. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (ACM Press, New York), 321-330.

Fredheim R, Moore A, Naughton J (2015) Anonymity and online commenting: The broken Windows effect and the end of drive-by commenting. *Proceedings of the ACM Web Science Conference* (ACM Press, New York), 11:1-8.

Freeman S, Grogger J, Sonstelie J (1996) The spatial concentration of crime. *Journal of Urban Economics* 40(2):216–231.

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.

Ghose A, Han S (2011) An empirical analysis of user content generation and usage behavior on the mobile Internet. *Management Science* 57(9):1671–1691.

Gneezy A, Imas A, Brown A, Nelson LD, Norton MI (2012) Paying to be nice: Consistency and costly prosocial behavior. *Management Science* 58(1):179-187.

Goes PB, Guo C, Lin M (2016) Do incentive hierarchies induce user effort? Evidence from an online knowledge exchange. *Information Systems Research* 27(3):497-516.

Goes PB, Lin M, Au Yeung CM (2014) "Popularity effect" in user-generated content: Evidence from online product reviews. *Information Systems Research* 25(2):222–238.

Gonzalez-Navarro M (2013) Deterrence and geographical externalities in auto theft. *American Economic Journal. Applied Economics* 5(4):92–110.
Griffiths TL, Steyvers M (2004) Finding scientific topics. *Proceedings of the National academy of Sciences* 101(1):5228-5235.
Hann IH, Roberts JA, Slaughter SA (2013) All are not equal: An examination of the economic returns to different forms of participation in open source software communities. *Information Systems Research* 24(3): 520–538.
Harkins SG, Petty RE (1982) Effects of task difficulty and task uniqueness on social loafing. *Journal of Personality and Social Psychology* 43(6):1214.
Hausman J, Hall BH, Griliches Z (1984) Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* 52(4):909-938.
Hausman C, Rapson DS (2018) Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics* 10:533-552.
Hoffman DL, Novak TP, Peralta MA (1999) Information privacy in the marketspace: Implications for the commercial uses of anonymity on the Web. *The Information Society* 15(2):129-139.
Huang N, Hong Y, Burtch G (2017) Social network integration and user content generation: Evidence from natural experiments. *MIS Quarterly* 41(4):1035–1058.
Huang P, Zhang Z (2016) Participation in open knowledge communities and job-hopping: Evidence from enterprise software. *MIS Quarterly* 40(3): 785-806.
Huang Y, Singh PV, Ghose A (2015) A structural model of employee behavioral dynamics in enterprise social media. *Management Science* 61(12):2825–2844.
Huberman BA, Adar E, Fine LR (2005) Valuating privacy. *IEEE Security & Privacy* 3(5):22-25.
Hwang EH, Singh PV, Argote L (2015) Knowledge sharing in online communities: Learning to cross geographic and hierarchical boundaries. *Organization Science* 26(6): 1593–1611.
Jabr W, Mookerjee R, Tan Y, Mookerjee V (2014) Leveraging philanthropic behavior for customer support: The case of user support forums. *MIS Quarterly* 38(1):187–208.
Jessup L, Connolly T, Galegher J (1990) The effects of anonymity on GDSS group process with an idea-generating task. *MIS Quarterly* 14(3):313-321.
Johnson S, Faraj S, Kudaravalli S (2014) Emergence of power laws in online communities: The role of social mechanisms and preferential attachment. *MIS Quarterly* 38(3):795–808.
Jones D, Linardi S (2014) Wallflowers: Experimental evidence of an aversion to standing out. *Management Science* 60 (7):1757-1771.
Kilner PG, Hoadley CM (2005) Anonymity options and professional participation in an online community of practice. *Proceedings of 2005 Conference on Computer Support for Collaborative Learning: Learning 2005: The Next 10 Years!* 272–280. International Society of the Learning Sciences.
Kling R, Lee Y, Teich A, Frankel MS (1999) Assessing anonymous communication on the internet: Policy deliberations. *Information Society* 15(2):79–90.
Kuang L, Huang N, Hong Y, Yan Z (2019) Spillover Effects of Financial Incentives on Non-Incentivized User Engagement: Evidence from an Online Knowledge Exchange Platform. *Journal of Management Information Systems* 36(1):289-320.
Lacetera N, Macis M, Slonim R (2012) Will there be blood? Incentives and substitution effects in pro-social behavior. *American Economic Journal: Economic Policy* 4(1):186–223.
Lacetera N, Macis M, Slonim R (2014) Rewarding volunteers: A field experiment. *Management Science* 60(5):1107–1129.
Lee YJ, Hosanagar K, Tan Y (2015) Do I follow my friends or the crowd? Information cascades in online movie ratings. *Management Science* 61(9):2241-2258.
Leshed G (2009) Silencing the clatter: Removing anonymity from a corporate online community. Online deliberation: Design, research, and practice 243-251.
Lowenthal PR, Dennen VP (2017) Social presence, identity, and online learning research: Research development and needs. *Distance Education* 38(2):137-140.

Miguel E, Kremer M (2004) Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72(1):159–217.

Miranda S, Saunders C (2003) The social construction of meaning: An alternative perspective on information sharing. *Information Systems Research* 14(1):87–106.

Nam H, Joshi YV, Kannan PK (2017) Harvesting brand information from social tags. *Journal of Marketing* 81(4):88-108.

Omernick E, Sood SO (2013) The impact of anonymity in online communities. *2013 International Conference on Social Computing*, 526–535.

Paskuda M, Lewkowicz M (2017) Anonymity interacting with participation on a Q&A site. *AI and Society* 32(3):369–381.

Png IP, Wang CY, Wang QH (2008) The deterrent and displacement effects of information security enforcement: International evidence. *Journal of Management Information Systems* 25(2):125-144.

Qiu L, Kumar S (2017) Understanding voluntary knowledge provision and content contribution through a social-media-based prediction market: A field experiment. *Information Systems Research* 28(3):529–546.

Roberts JA, Hann IH, Slaughter SA (2006) Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the apache projects. *Management Science* 52(7): 984–999.

Scott SV, Orlikowski WJ (2014) Entanglements in practice: Performing anonymity through social media. *MIS Quarterly* 38(3):873-893.

Shen W, Hu JY, Ulmer JR (2015) Competing for attention: An empirical study of online reviewers' strategic behavior. *MIS Quarterly* 39(3):683–696.

Short J A, Williams E, Christie B (1976) The social psychology of telecommunications (John Wiley & Sons, Hoboken, NJ).

Sia CL, Tan BCY, Wei KK (2002) Group polarization and computer-mediated communication: Effects of communication cues, social presence, and anonymity. *Information Systems Research* 13(1):70–90.

Singh PV, Sahoo N, Mukhopadhyay T (2014) How to attract and retain readers in enterprise blogging? *Information Systems Research* 25(1):35–52.

Soetevent AR (2005) Anonymity in giving in a natural context—a field experiment in 30 churches. *Journal of Public Economics* 89(11):2301–2323.

Sun B (2005) Promotion effect on endogenous consumption. *Marketing Science* 24(3):430–443.

Sun M, Zhu F (2013) Ad revenue and content commercialization: Evidence from blogs. *Management Science* 59(10):2314–2331.

Toubia O, Stephen AT (2013) Intrinsic vs. image-related utility in social media: Why do people contribute content to Twitter? *Marketing Science* 32(3):368–392.

Wang AS, Pang MS, Pavlou P (2017) Cure or poison? Impact of identity verification on the creation of fake posts on social media. In *ICIS 2017 Proceedings*. AIS.

Wasko MM, Faraj S (2005) Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly* 29(1):35–57.

Zhang XM, Zhu F (2011) Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review* 101(4):1601–1615.