Computers in Human Behavior 127 (2022) 107042

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Engaging voluntary contributions in online review platforms: The effects of a hierarchical badges system

Dandan Ma [a], Shuqing Li [a], Jia Tina Du [b,*], Zhan Bu [a], Jie Cao [a], Jianjun Sun [c]

[a] School of Information Engineering, Nanjing University of Finance and Economics, Nanjing, China
[b] UniSA STEM, University of South Australia, Adelaide, South Australia, Australia
[c] School of Information Management, Nanjing University, Nanjing, China

ARTICLE INFO

Keywords:
Hierarchical badges system
Online review
Incentive mechanism
User effort
Opinion

ABSTRACT

Employing a hierarchical badges system to sustain users' participation and content contribution has been a common practice in online review platforms. However, it is still unclear whether this incentive mechanism is effective in inducing users to provide more substantial and useful reviews. In this paper, we draw on the goal-setting theory and status hierarchies theory to study how the hierarchical badges system impacts users' behavior such as their effort levels and expressed opinions, using the data of 234,220 reviews from a large online review platform. Our results show that users significantly increase their efforts as their status are promoted, manifesting as review frequency, length, and readability. In terms of expressed opinions, users with higher badges are less likely to post extremely negative ratings and use negative emotional words. Our results have important implications for business modes that rely on user-generated content, such as online review platforms, social question and answer sites, social media, and crowdsourcing systems.

1. Introduction

Individuals are increasingly relying on online reviews to make decisions about which restaurants, hotels, movies, books, and news are worth their time and money. Online reviews contain rich information on product attributes and qualities as experienced by consumers, which can alleviate information asymmetry in the online market (Burtch et al., 2018). Online review platforms such as Yelp, TripAdvisor, and Rotten Tomatoes, achieve great success in the past few decades. Such platforms critically depend on the ability to continuously induce quality reviews contributed by users (Goes et al., 2016). Unfortunately, online reviews often suffer from an undersupply problem, due to its public good nature (Gallus, 2017; Chen et al., 2018). Like many user-generated contents (UGCs), online reviews are typically provided on a voluntary basis, making it difficult for platforms to guarantee its volume and quality. As "1% rule" shows, only 1% of the users of a platform use reviews and actively contribute it, while the other 99% of the users just lurk (Huang et al., 2018). Furthermore, when users do provide reviews, they are often brief and susceptible to certain biases (Muchnik et al., 2013; Bu et al., 2019), limiting their helpfulness to other consumers (Cao et al., 2011).

Encouraging users to contribute reviews is thus an important issue for online review platforms. To address this problem, many platforms have been experimenting with various types of intervention strategies, e.g., Amazon provides free or discounted sample products to write product reviews; Yelp awards users "elite" badge to signal their contributions. One intervention that has received little consideration in existing literature is the hierarchical badges system, where users are afforded different levels of badge according to their contributions (Ma et al., 2020). A hierarchical badges system consists of progressively more challenging goals and increasingly higher status (Anderson et al., 2013). Regularly, in a hierarchical badges system, users may receive virtual points by making contributions and an achievement badge will be rewarded to them in front of their peers as their cumulative points reach higher thresholds. The badges are not binary but often hierarchical, and it will be progressively more difficult to increasingly achieve a higher badge (Goes et al., 2016; Wang, Lu, & Tan, 2018). These achievement badges are extrinsic, nonmonetary motivators. Each badge represents a user's social status and reputation in an online community. A more prestigious badge signals a more recognized, experienced, and influential user (Antin & Churchill, 2011). At present, such hierarchical badges system has been implemented by an increasing number of online review

* Corresponding author.
E-mail address: tina.du@unisa.edu.au (J.T. Du).
URL: http://people.unisa.edu.au/tina.du (J.T. Du).

https://doi.org/10.1016/j.chb.2021.107042
Received 1 March 2021; Received in revised form 28 August 2021; Accepted 30 September 2021
Available online 7 October 2021

platforms, with the hope of stimulating more and useful reviews. Despite its popularity in practice, however, relatively little attention has been given to whether this incentive mechanism is effective in inducing reviews (Liu et al., 2016; Lu et al., 2019).

In this paper, we fill this gap by seeking to understand the role played by a hierarchical badges system on users' review behavior. Drawing on goal-setting theory and status hierarchies theory, we study whether the hierarchical badges system induces users to increase their contributions (e.g., frequency of reviewing, length of reviews, readability of reviews) to the online community. And, we investigate if users' opinions (e.g., ratings and the sentiments of review texts) change as their status are promoted. Specifically, we address the following two research questions:

(1) How does a reviewer's status rank in a hierarchical badges system affect his or her *effort level*, including review frequency, review length, and review readability ?
(2) How does a reviewer's status rank in a hierarchical badges system affect his or her *opinions expressed*, including average rating, ratio of extremely negative ratings, and review sentiment?

The rest of the paper proceeds as follows. We review related literature and develop hypotheses in section 2. Details of research design including data collection, variables measurement and model specification are described in section 3. Section 4 reported the empirical results and offers an interpretation of the findings. Theoretical and practical implications are discussed in section 5. Section 6 concludes the paper by discussing the limitations and identifying a number of avenues for future work.

2. Related literature and hypotheses development

2.1. Related literature

2.1.1. Incentive mechanisms in UGC platforms

Our study mainly builds on and contributes to the field of incentive mechanisms in UGC platforms. To encourage users to contribute more contents, UGC platforms have deployed various incentive mechanisms, which can be divided into financial incentives such as a small cash (Stephen et al., 2012), free products (Fayazi et al., 2015; Qiao et al., 2020), and rebates (Cabral & Li, 2015), or gamified incentives such as badges, points, and levels (Anderson et al., 2013; Hamari, 2017; Goes et al., 2016; Hanson, Jiang, & Dahl, 2019).

2.1.1.1. Financial incentives. Offering financial incentives is a natural way to motivate desired actions. It has attracted much attention from academic communities where they seek to explore whether financial incentives are effective for motivating users to contribute actively. Empirical studies report contradictory results regarding the impact of financial incentives. On the one hand, providing financial incentives will benefit the platform. It can motivate users to generate more reviews (Burtch et al., 2018; Wang et al., 2019) and enrich the diversity of topics involved in the reviews (Qiao et al., 2020). On the other hand, it is found that paying for reviews will lead to unexpected negative effects. First, it may suppress users' intrinsic motives to contribute (Deci et al., 1999), leading consumers to exert less effort in writing a review. For example, Qiao et al. (2020) conducted a field experiment on Amazon and found that financial incentives result in a decrease in efforts as reflected in the usefulness and the lexical richness of reviews. Second, financial incentives may bias users' opinions. Cabral and Li (2015) examined the effect of offering a rebate in exchange of a review on eBay. They found that offering rebates significantly decreases the probability of providing negative feedback. Another example is Khern-am-nuai et al. (2018)'s study where they found users tend to write positive reviews after the implementation of monetary incentives. Third, consumers tend to doubt the authenticity of reviews and make negative inferences about product qualities if they learn that the review was paid (Stephen et al., 2012).

2.1.1.2. Gamified incentives. Gamified incentives are increasingly used as an alternative method to encourage user contribution in many UGC platforms, as financial incentives are costly and have aforementioned shortcomings. Gamified incentives refer to applying elements drawn from game designs into non-game context to improve individuals' motivation and engagement in a task (Liu et al., 2017). Common game design elements include points, badges, levels, and leaderboards (Hamari et al., 2014). Gamified incentives have been widely used in various contexts for diverse proposes, including improving working performance (Li et al., 2012), encouraging people to live healthy (Keung et al., 2013), motivating students to engage in e-learning (Su & Cheng, 2015), and increasing user engagement in software systems (Darejeh & Salim, 2016). Despite its popularity in practice, however, a rather limited amount of research has empirically examined the effectiveness of gamified incentives in the context of UGC platforms (Hamari, 2017). And, a rich array of psychological mechanisms enlightening why and how gamified incentives affect user contribution behavior remains unexplored (Liu et al., 2017).

As far as we know, only several studies have attempted to empirically explore the effect of gamified incentives on user contribution ((Hamari, 2017); Goes et al., 2016; Li et al., 2019). Gamified incentives were found to have a positive impact on user contribution while its effect depends on the context in which the game design elements are implemented (Hamari et al., 2014). For example, in an online peer-to-peer trading platform, Hamari et al. (2017) found the implementation of a game design element called 'badges' positively affected user activity. After the implementation, users were significantly more likely to post trade proposals, carry out transactions, comment on proposals and use the platform. In an online question and answering community, Goes et al. (2016) examined the effect of hierarchical badges systems on users' contribution frequency. They found a transiently positive effect where users increase their contribution frequency before the badges are reached, but significantly reduce it after that. In an online review platform, Li et al. (2019) studied how awarding users an "elite" badge affects the linguistic characteristics of online reviews they post. Their results showed that elite users exhibit more analytical thinking, positive emotional tone, stronger authenticity, and stronger clout in their reviews than non-elite reviewers.

In summary, while prior studies have shown several consequences of gamified incentives, there remains a paucity of a coherent and ample body of empirical evidence on its effectiveness. Our study aims to fill this gap by examining how a gamified incentive mechanism called "hierarchical badges system" affects user contribution behavior in the context of an online review platform. In order to better understand how hierarchical badges influence user contribution behavior in terms of their efforts and opinions, we draw on goal-setting theory and status hierarchies theory.

2.1.2. Theoretical underpinnings

2.1.2.1. Goal-setting theory. As suggested by Goes et al. (2016), each badge level in the hierarchical badges system represents a predefined goal that users pursue. Thus, the goal-setting theory in psychology can serve as an appropriate theoretical underpinning for us to investigate the effect of hierarchical badges on user contribution behavior. Goal-setting theory has been studied for decades to explain how to motivate people to perform better in organizations by setting and managing goals (Locke, 1967, 1968). This theory states that setting goals for individuals leads to higher performance compared with no goals. Particularly, this effect is further strengthened if the goals are challenging and specific (Locke et al., 1981). For example, challenging goals are better than too difficult or easy goals; specific and quantifiable goals are better than "do your best" goals. Drèze and Nunes (2011) examined the effect of successfully attaining a goal on future effort to achieve the next goal in a recurring goals framework. This study is particularly relevant since the hierarchical badges system is essentially a recurring goals framework in which goals are represented in the form of badges (Gutt et al., 2020). Drèze and Nunes (2011) found that successful goal achievement has a positive effect on subsequent user effort, but only if the goals are challenging. Self-efficacy was identified as an important psychological mechanism for this effect. According to Bandura (1993), self-efficacy refers to a person's belief in his or her ability and capacity to accomplish a task. Both assigning goals and completion of goals can enhance individual self-efficacy, thus increasing users' efforts to achieve the next goal.

Goal-setting theory has been used to explain user contribution behavior in online communities in prior studies. For instance, Goes et al. (2016) examined users' efforts (measured by contribution frequency) before and after badge attainment in an online question and answering community. They have found that users increase their efforts before badge attainment, but reduce their efforts once they reach a badge. Tondello et al. (2018) proposed a conceptual framework to explain how gamification incentives in online communities work through the lens of goal-setting theory. Gutt et al. (2020) examined how goals represented by badges affect user efforts (measured by contribution frequency) in an online question and answering community. The results provide evidence that goal difficulty mediate the positive effect of successful goal attainment on subsequent user effort. In line with these studies, goal-setting theory lays the theoretical basis for us to investigate the effect of hierarchical badges on user effort.

2.1.2.2. Status and status hierarchies. Sociopsychologists have found that status is a fundamental motivating force of human activity (Anderson et al., 2015). Status was defined as an individual's relative standing in a group as determined by respect and deference (Magee & Galinsky, 2008). It includes two major components. First, status involves *respect and deference*. Individuals with high status are afforded regard, esteem, recognition, attention, or importance by peers. Second, status involves *relative standing*, in that individuals are rank-ordered and positioned in a hierarchy. Status is directly related to the hierarchical badges system we study, because each level of badge signals a user's status and rank in the community (Antin & Churchill, 2011). Studies have shown that individuals pay close attention to status, not only in physical organizations but in online communities (Gallus, 2017; Levina & Arriaga, 2014). Even with high fluidity and anonymity, online communities are still organizations of people (Goes et al., 2016). Hence, it is reasonable to relate the research of status and status hierarchies in organizations to the hierarchical badges system in online communities.

It is established in sociopsychology that people's cognition and behavior change when they receive different status in a hierarchy (Staw, 1976; Willer, 2009). Willer (2009) claimed that higher-status individuals valued the group interests more, while lower-status individuals seemed to be more self-interested. With the improvement of status, individuals are more inclined to view the group positively and relate the group interests to their own. This awareness will cause increased pro-social behaviors such as greater giving in groups (Dong et al., 2020). Staw (1976) found that higher-status individuals were more cautious with their established reputation, so they were more conservative and decreasing in risk-seeking. Zhang et al. (2020) proposed that the reason for the higher-status users to show socially desirable behaviors in online communities is that they want to maintain the benefits of being high status. Individuals tend to seek higher status in a hierarchical organization because it can help individuals access greater economic or social resources (Perretti & Negro, 2006). Resources are unequally distributed across status hierarchies where individuals with higher status possess more resources (Anderson & Brion, 2014). In online communities, a higher status can help individuals access more economic or social benefits, such as the chance of receiving free products, or more respect by other community members. Thus, status can serve as a driver for certain behaviors, especially those that can help users achieve higher status (Goes et al., 2016).

2.2. Hypotheses development

This paper aims to study how users' efforts and opinions respond to hierarchical badges. Therefore, in this section, we developed research hypotheses from two aspects: 1) review frequency, length and readability which reveal users' efforts in contributing reviews; 2) average rating, ratio of extremely negative ratings, and review sentiment which reveal users' opinions.

2.2.1. Hierarchical badges and user effort

Two distinct mechanisms suggest that hierarchical badges could have a positive effect on user effort. First, hierarchical badges can serve as goals users pursue, thus motivating users' efforts (Goes et al., 2016; Gutt et al., 2020). It is well established in goal-setting literature that setting goals can improve individual performance especially if the goals are specific and challenging (Locke et al., 1981). In a hierarchical badges system, the points thresholds of different levels of badges were clearly defined and it is usually more challenging to get a higher badge. Thus, in line with previous study on goal-setting theory, we would expect to see users increase their efforts as they pursue a higher badge. Second, hierarchical badges represent a user's status in online communities. According to status hierarchies theory, hierarchical badges will incentivize an individual behaving more pro-social behavior, such as making contributions to the community, because they want to access the benefits of being high status (Dong et al., 2020; Perretti & Negro, 2006). Thus, within a status hierarchy such as a hierarchical badges system, we would also expect that users increase their efforts as they reach a higher status. Typically, user effort or user contribution was reflected by the number of reviews and the quality of reviews (Burtch et al., 2018; Goes et al., 2016; Zhang et al., 2020). Review quality is generally measured by review length and readability, which are two indicators of writing quality (Ghose & Ipeirotis, 2011). Based on the above analysis, it is supposed that users are more likely to post more, longer, and easier to read reviews as they devote more effort in writing reviews. Therefore, we propose the following hypothesis:

H1. Reviewers will invest more effort in their reviews as their badge upgrade, resulting in higher review frequency, length, and readability (see Fig. 1).

2.2.2. Hierarchical badges and user opinion

According to status hierarchies theory, the opinions users express may also be affected by a user's status rank in the hierarchy. In the

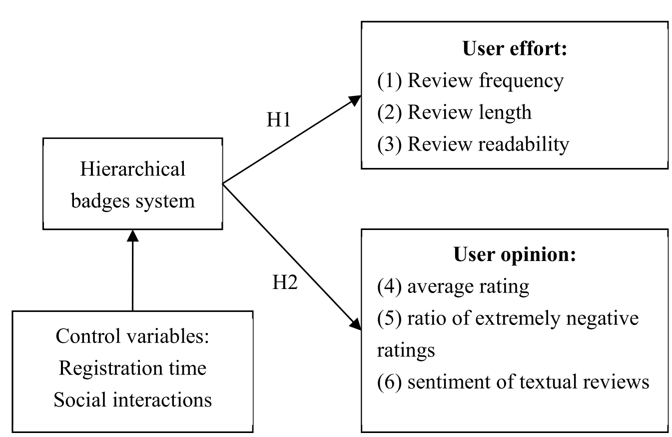

Fig. 1. A conceptual model of the effects of a hierarchical badges system on user behavior.

context of online review platforms, users' opinions toward products or services are normally expressed in two forms, e.g., the numeric ratings and textual review. Numerical ratings typically range from 1 to 5, where 1 and 5 indicate extremely negative and extremely positive, respectively, and the other ratings (2, 3, and 4) indicate moderate ratings (Mudambi & Schuff, 2010). To quantify numerical ratings, a natural metric that we are using is the average rating. Status hierarchies literature showed that individuals of high status are less likely to develop negative attitudes and are more prone to express positive emotions in social settings than those of low status (Kemper, 1991). Thus, as users' status within the community upgrade, we expect them to post more positive opinions, which is reflected by higher average rating. A second dimension of numerical ratings that we are proposing is the ratio of extremely negative ratings. Prior studies have shown that extremely negative ratings are more influential than moderate ratings and extremely positive ratings (Mudambi & Schuff, 2010; Park & Nicolau, 2015). Thus, it is vital to explore who are more likely to post extremely negative ratings and to what extent such choice is affected by status hierarchy. As extremely negative ratings are more likely to be imperial and risk-seeking than moderate ratings (Zhang et al., 2020), we hypothesize that high-status users are less likely to post extremely negative ratings compared with those low-status users. Moreover, the importance of mining the sentiments expressed in textual reviews has been acknowledged in many studies. For example, Ghose and Ipeirotis (2011) reported that numeric ratings cannot fully capture the polarity information in the review. Hu et al. (2014) found that textual reviews contain rich information on reviewers' feelings, experiences, and emotions about the product or service, beyond the numeric ratings. Therefore, we are also paying attention to the opinions expressed in textual review. As mentioned earlier, higher-status users have more to lose and are thus less motivated to take risk that dampen their established status (Staw, 1976). Hence, we calculate the sentiment scores of each textual review by sentiment analysis and hypothesize that as users' status upgrade, they are likely to use fewer negative emotional words in their textual reviews to mitigate risks, resulting in higher sentiment scores. We hence propose the following hypothesis:

H2. Reviewers will less likely express extremely negative opinions in their reviews as their badge upgrade, resulting in higher average rating, lower ratio of extremely negative ratings, and higher sentiment scores of textual reviews (see Fig. 1).

3. Research design

3.1. Data collection

To empirically test the above hypotheses, we collected the reviewers' information from Dianping.[1] Dianping is a leading Chinese review community, with more than 650 million users and 150 million reviews covering over 20 million merchants. The data were collected in December 2019. Since it is impractical to retrieve all reviewers' information across the website, we focus on these from certain products. First, we identified a target product which has gained large amounts of reviews, and its average rating should be neither too high nor too low. A too high/low average rating means the product in some extent attracts users tending to post high/low ratings, which may lead to a sample selection bias. As a result, we chose a product with 2093 reviews and an average 3-star ratings (Ratings range from 1 to 5 stars). We then retrieved users' profiles who posted reviews for the target product by using a python-based crawler. Since each review is provided by a different user, 2093 users' profiles were retrieved. For each user, we collected his/her badge level, contribution points, the registration time, number of followers, number of social interactions,[2] and most importantly all review information including numerical ratings and review texts. Furthermore, 426 users whose ratings or review texts are not complete were removed. Therefore, our final data set contains the information of 1667 users and the 234,220 reviews they posted.

3.2. Variables and measurement

With the dataset, we calculated some variables to measure users' badge levels, efforts, and opinions.

3.2.1. Badge levels

To explore the impact of hierarchical badges system on users' efforts and opinions, users' badge levels were employed as the key independent variable. Dianping adopted an eight-level badges system in terms of the difficulty to gain them, as shown in Table 1. Users receive points by posting reviews, and once a sufficient number of points are accumulated, users' rank on the platform will be elevated to a higher badge level. The thresholds of points for higher badges increase exponentially. For instance, users who post one review will gain one-star badge, two-star badge requires 100 points which means 5 reviews, and the three-star badge requires 300 points, four-star requires 1000 points, and so on. Each level of badge has a symbolically logo, which obviously displays on users' profile to signal users' status. The badge logo also appears next to the username and avatar in each review. For empirical analysis, we used numerical equivalent values (1–8) to represent the badge levels.

3.2.2. User effort

In this study, we define user effort as the time and energies that users spend in writing reviews. In prior studies, user effort was regularly measured by contribution frequency (Goes et al., 2016; Gutt et al., 2020), review length (Burth et al., 2018, Sun, Dong, & McIntyre, 2017), review helpfulness (Liu et al., 2016; Sun et al., 2017), lexical richness (Qiao et al., 2020) of reviews and review readability (Goes et al., 2014). As shown in Table 2, three variables that reveal users' efforts of contributing reviews were extracted from the data set, namely review

Table 1
Hierarchical badges system in Dianping.

Numerical equivalent values	Badge levels	Badge logos	Contribution points
1	One-star		below 100
2	Two-star		101–300
3	Three-star		301–1000
4	Four-star		1001–3000
5	Quasi five-star		3001–8000
6	Five-star		8001–15000
7	One-diamond		15001-30000
8	Two-diamond		Above 30000

[1] https://www.dianping.com/.

[2] The online review platform in our research has the function of social interaction where users can easily connect with each other. For instance, a user may follow other users, like other users' reviews, rate others' reviews as helpful or unhelpful, or comment on others' reviews. The platform measures social interactions by using the number of followers a user has, and the interactions with other users in terms of votes, likes and comments.

Table 2
Variables used for analysis.

Types of variables	Variables	Definition
Independent variable	Badge levels	The badge level of a user on Dianping (see Table 1), measured by the numerical equivalent values 1–8.
Dependent variables-User effort	Review frequency	Average number of reviews posted by a user per month, measured by the number of reviews of each user/the number of months since registration of each user.
	Review length	Average number of words per review per user, measured by the number of words contained in reviews of each user/the number of reviews of each user.
	Review readability: GFI	The degree of difficulty for readers to understand a given text, measured by 0.4* (the average sentence length + the ratio of hard words *100).
Dependent variables-User opinion	Average rating	The average of each user's ratings, measured by total ratings/the number of ratings of each user.
	Ratio of extremely negative ratings	The ratio of 1-star ratings of each user, measured by the number of 1-star ratings/the number of ratings of each user.
	Review sentiment	The positive or negative sentiment tendency expressed in review texts, measured by snowNLP and the value range is 0–1.
Control variables	Social interactions	Number of interactions of each user, measured by Dianping.
	Registration time	Number of months a user has been on Dianping since registration.

frequency, length, and readability. First, review frequency was measured by the average number of reviews per month since a user registered. Second, review length was calculated by counting the number of words in each review. Third, review readability was measured by a popular metric: the Gunning Fog Index (GFI), which has been widely used by prior studies on online reviews (Ghose & Ipeirotis, 2011; Goes et al., 2014; Liu et al., 2018). This metric is calculated by two parts: the average sentence length and the ratio of hard words in each text. The calculation formula is as follows.

$$GFI = 0.4*\left(\frac{number\ of\ words}{number\ of\ sentences} + \frac{number\ of\ hard\ words}{number\ of\ words} * 100\right) \quad (1)$$

The GFI measures how easy it is for readers to understand a given text and is inversely correlated with review readability. A higher GFI means that a text contains more hard words and longer sentences, and is difficult to understand. In Gunning (1969)'s original work, the hard words refer to the words with more than two syllables for each 100 words in English text. In this study, we use the Syllabus of Graded Vocabulary for Chinese Proficiency to identify the hard words in Chinese text. We developed a python-based program to calculate the GFI of each review according to model (1).

3.2.3. *User opinion*

Three variables that reflect users' opinions were extracted from the data set, namely average rating, ratio of extremely negative ratings, and review sentiment (Table 2). Average rating is the average of all the ratings of each user, which has been normalized to the range of 0–1. Ratio of extremely negative ratings refers to the ratio of 1-star in the total number of ratings per user. This study uses review sentiment to capture the reviewers' textual opinions. Review sentiment refers to the sentiment tendency expressed in review texts, both positive and negative. Sentiment analysis is a common textual data mining method that can analyze the sentiment tendency of a text. Specifically, we adopt snowNLP, a python kit that specializes in sentiment analysis of Chinese texts, to carry on sentiment analysis. SnowNLP is free for academic research and has been tested and validated in prior studies (Tseng et al., 2018; Wang et al., 2018). It can predict the probability that a review text is positive or negative, and reports a sentiment score between 0 and 1 for each review text, where 0 denotes extremely negative and 1 means extremely positive.

3.3. *Empirical models*

As mentioned above, we are interested in the efforts and opinions of users with different badge levels. Therefore, badge level is the key independent variable in our model. The six metrics that reflect reviewers' efforts and opinions are dependent variables: (1) *review frequency, (2) review length, (3) review readability: GFI, (4) average rating, (5) ratio of extremely negative ratings, and (6) review sentiment.* To exclude potential exogenous factors, we also include some control variables: the length of time a reviewer registered on Dianping, and the number of social interactions a reviewer has. Prior empirical studies have demonstrated that these two variables also have effect on user efforts and opinions (e.g., Goes et al., 2014; Zhang et al., 2020). However, these two variables are not of primary interest to our study. They should be controlled in model in order to eliminate their potential effect on the dependent variables (user efforts and opinions), thus allowing for a more accurate detection of the effects of the independent variables of interest (badge level).

To test the H1and H2, we used ordinary least squares (OLS) regression, and specified our model as follows:

$$DV_i = \alpha_0 + \alpha_1 Badge\_level_i + \alpha_2 Registration\_time_i + \alpha_3 Social\_interactions_i + \varepsilon_i \quad (2)$$

Where DV_i represents the six dependent variables that reflect efforts and opinions of *reviewer*$_i$, *Badge_level*$_i$ refers to the independent variable of *reviewer*$_i$'s badge level. *Registration_time*$_i$ and *Social_interactions*$_i$ are two control variables that represent the length of time *reviewer*$_i$ have registered on Dianping, and the number of social interactions he or she has respectively. In this formula, we are particularly interested in the coefficient α_1, which estimates the effect of badge level on users' efforts and opinions.

4. Results

4.1. *Descriptive statistics*

Table 3 presents the descriptive statistics of the main variables used in our model. The badge level for each user ranges from 1 to 8, with an average of 4.9. In terms of users' efforts, a user posts around 2 reviews on average per month, and the average length for each review is 135 words. The review readability, which is measured by GFI, ranges from about 2.67 to 61.16, with an average of approximately 8. In terms of the

Table 3
Descriptive statistics of the dataset.

Variables	N	Mean	Std. Deviation	Minimum	Maximum
Badge level	1667	4.90	1.54	1	8
Review frequency	1667	1.98	4.20	0.01	64.03
Review length	1667	135.29	92.99	15	781.13
Review readability: GFI	1667	8.01	2.55	2.67	61.16
Average rating	1667	4.42	0.53	1	5
Ratio of extremely negative ratings	1667	0.03	0.09	0	1
Review sentiment	1667	0.90	0.30	0	1
Social interactions	1667	1164.94	9297.58	0	303424
Registration time	1667	73.07	31.12	6.97	193.53

users' opinions, the average rating is 4.42, the ratio of extremely negative ratings is 3% on average, and the average review sentiment contained in each textual review (measured by snowNLP with a value range 0–1) is 0.9. These figures suggest that online reviews tend to be overwhelmingly positive. The average number of social interactions for each user is about 1165, the length of time for each user has been on Dianping ranges from about 7 months to 194 months, with an average of around 73 months.

4.2. Exploratory analysis

Before testing the hypotheses, we first conduct some exploratory analysis on our dataset to obtain a preliminary view on how users' efforts and opinions change as their badges upgrade. Specifically, we set badge level as the axis-X and plot the dynamic of users' efforts and opinions in Fig. 2. Z-score normalization method was used to plot the different dependent variables in one figure. Fig. 2 (a) shows the relationship between users' badge levels and their efforts. We can see that review frequency and review length generally show a steady trend of increasing with the badge levels. Namely, as the badge level grows, users post more reviews, and probably longer ones. The review readability metric: GFI, generally decreases as a user gets higher badges. It is noted that GFI is inversely correlated with readability, which suggests that review readability increases as a user's status is promoted. In summary, the result to some extent is in line with our hypotheses, suggesting that hierarchical incentives can motivate users to make more efforts, i.e., to post more, longer, and easier to read reviews.

Fig. 2 (b) reveals how users' opinions change as their badge levels upgrade. We find that the average rating of each user is above 0.82 (equivalent to 4.1 score) and up to 0.92 (equivalent to 4.6-score), which suggests that users tend to post positive opinions regardless of their badge levels. The sentiment analysis of textual reviews shows a similar trend as the average ratings. It is noteworthy that the ratio of extremely negative ratings (1-score) is continually declining from badge level 2, which roughly indicates that users with higher badges are less likely to post extremely negative ratings. Fig. 2 provides model-free evidence that most of the variables roughly change in line with our hypotheses.

Furthermore, to explore the collinearity problem among variables, Table 4 presents the Pearson correlation coefficients among them. The Pearson correlation coefficients of all variables are less than 0.3, suggesting that there is no obvious collinearity between the variables. In addition, as shown in the first column in Table 4, the badge level is significantly correlative with five of the six dependent variables, namely, review frequency, review length, review readability, ratio of extremely negative ratings, and review sentiment. Therefore, Table 4 provides the possibility for further hypotheses testing.

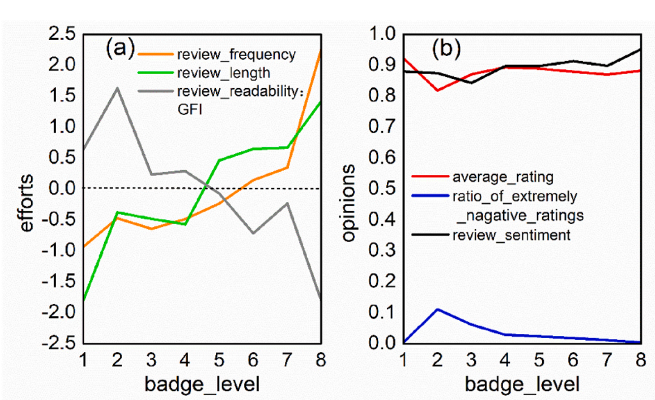

Fig. 2. Users' efforts (a) and opinions (b) of different badge levels.

4.3. Hypotheses testing

4.3.1. Badge level and user effort

In order to test H1, we conduct OLS regression on the basis of model (2) for the three dependent variables of user effort (i.e., review frequency, length, and readability). To alleviate the potential sampling bias, we also conduct bootstrapped OLS. Table 5 presents the regression coefficients and the standard errors in parentheses for OLS and bootstrapped OLS regression. As the results of bootstrapped OLS regression show consistent trends with OLS, our analysis is mainly based on OLS results. As seen in the first row, all the coefficients of the key independent variable badge level are significant and consistent with our hypotheses, suggesting users increase their review frequency, length, and readability as their badges upgrade. Specifically, we can see the coefficient on review frequency is 0.585, significant at the 0.001 level, indicating that a one-level upgrade can induce 0.585 increase in the number of review frequency. Regarding review length, the coefficient is 9.651 and significant at 0.001 level, indicating that a one-level upgrade can induce 9.651 increase in review length. As for review readability: GFI, the coefficient is −0.159 and significant at 0.001, indicating that a one-level upgrade can induce 0.159 increase in review readability. These figures provide strong evidence that the hierarchical badges system has a positive effect on user effort. The review frequency, length, and readability all significantly increase as users' badge levels upgrade, thus supporting H1.

In addition, the coefficients of social interactions are only weakly significant on review frequency (see the second row of Table 5), but not review length and readability, indicating that users with more social interactions tend to post more reviews per month. The coefficients of registration time on review frequency, length and readability are all significant (see the third row of Table 5). Specifically, registration time has a negative effect on review frequency, length, and GFI (the coefficient is −0.007, −0.267 and −0.014 respectively), indicating that older members are less likely to post reviews, and their reviews are shorter but easier to read.

4.3.2. Badge level and user opinion

Table 6 shows the regression results based on model (2) for the three dependent variables of user opinion (i.e., average rating, ratio of extremely negative ratings and review sentiment). Similarly, the results of bootstrapped OLS regression are qualitatively consistent with OLS. As seen in the first row, when the dependent variable is average rating, the coefficient is statistically nonsignificant, whereas the coefficient is statistically significant when estimating the effect on ratio of extremely negative ratings and review sentiment. The possible reason why badge level has no significant effect on average rating may be related to the feature of overwhelmingly positive of online rating. As shown in Fig. 2 (b), reviewers tend to post positive reviews regardless of their badge level. Hence, the average rating can't lend support for H2. However, using the ratio of extremely negative ratings and review sentiment as indicators of user effort, we find evidence in support of H2. Specifically, the coefficient on ratio of extremely negative ratings is negative (−0.010) and significant at 0.001, indicating that a one-level upgrade can induce 1% decrease in the ratio of extremely negative ratings. Moreover, the coefficient of badge level on review sentiment is positive (0.011) and significant at 0.05, indicating that a one-level upgrade can induce 1.1% increase in review sentiment. Both results show that reviewers tend to post fewer extremely negative ratings and likely write more positive words in textual reviews as their badge levels upgrade, providing strong evidence for H2. Hence, H2 is partially supported by the dependent variables of ratio of extremely negative ratings and review sentiment.

The regression results of two control variables were also presented in Table 6. As seen in the second row, the coefficients of social interactions on three indicators of user opinion are all statistically nonsignificant, indicating that social interactions have no evident effect on user

Table 4
Correlation among variables.

	Badge level	Review frequency	Review length	Review readability: GFI	Average rating	Ratio of extremely negative ratings	Review sentiment	Social interactions	Registration time
Badge level	1								
Review frequency	0.218**	1							
Review length	0.148**	0.196**	1						
Review readability: GFI	−0.114**	−0.282**	−0.043	1					
Average rating	−0.001	0.003	−0.009	−0.006	1				
Ratio of extremely negative ratings	−0.169**	−0.060*	−0.065**	−0.002	−0.648**	1			
Review sentiment	0.060*	0.095**	−0.002	−0.053*	0.132**	−0.121**	1		
Social interactions	0.223**	0.092**	0.027	−0.025	−0.007	−0.032	0.032	1	
Registration time	0.102**	−0.031	−0.073**	−0.177**	−0.047	−0.015	−0.007	−0.030	1

Note: ** denotes correlation is significant at the 0.01 level (2-tailed); * denotes correlation is significant at the 0.05 level (2-tailed).

Table 5
Effect of badge level on user effort.

	Review frequency		Review length		Review readability: GFI	
	OLS	Bootstrapped OLS	OLS	Bootstrapped OLS	OLS	Bootstrapped OLS
Badge level	0.585*** (0.068)	0.568*** (0.100)	9.651*** (1.513)	9.575*** (1.498)	−0.159*** (0.041)	−0.157** (0.050)
Social interactions	0.000+ (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Registration time	−0.007* (0.003)	−0.007+ (0.004)	−0.267*** (0.073)	−0.269*** (0.068)	−0.014*** (0.002)	−0.014*** (0.002)
Constant	−0.405 (0.396)	−0.342 (0.343)	107.659*** (8.870)	108.154*** (8.496)	9.795*** (0.242)	9.791*** (0.224)
F	30.408***		16.977***		23.623***	
N	1667		1667		1667	

Note: *** denotes regression coefficients is significant at the 0.001 level; ** denotes regression coefficients is significant at the 0.01 level; * denotes regression coefficients is significant at the 0.05 level;+denotes regression coefficients is significant at the 0.1 level.

Table 6
Effect of badge level on user opinion.

	Average rating		Ratio of extremely negative ratings		Review sentiment	
	OLS	Bootstrapped OLS	OLS	Bootstrapped OLS	OLS	Bootstrapped OLS
Badge level	0.002 (0.009)	0.002 (0.010)	−0.010*** (0.001)	−0.010*** (0.002)	0.011* (0.005)	0.011* (0.005)
Social interactions	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Registration time	−0.001* (0.000)	−0.001* (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	4.468*** (0.052)	4.469*** (0.064)	0.076*** (0.008)	0.076*** (0.013)	0.849*** (0.029)	0.850*** (0.031)
F	1.312		16.253***		2.295+	
N	1667		1667		1667	

Note: *** denotes regression coefficients is significant at the 0.001 level; ** denotes regression coefficients are significant at the 0.01 level; * denotes regression coefficients is significant at the 0.05 level;+denotes regression coefficients is significant at the 0.1 level.

opinion. As for registration time, the coefficient is only significant on average rating, suggesting that older members tend to post fewer positive ratings.

5. Conclusions and implications

Hierarchical badges system is an increasingly widespread feature of the UGC community, with expecting that it will elicit significant incentive effects on user behavior. Yet there is a research gap in our understanding of how such a hierarchical incentive mechanism actually affects users' behaviors (Liu et al., 2016; Lu et al., 2019). In this paper, we have attempted to provide more systematic empirical evidence for this important but underdeveloped question. Drawing from goal-setting and status hierarchies theories, we proposed two hypotheses to explore these effects. Particularly, six dependent variables were developed to capture reviewers' efforts and opinions, that is, review frequency, review length, review readability: GFI, average rating, ratio of extremely negative ratings, and review sentiment. Based on the data collected from Dianping online community, we found evidence for the proposed hypotheses. Our results indicate that the hierarchical badges system has a significant effect on reviewers' contributions in terms of: 1) reviewers' efforts including the review frequency, length and readability significantly increase with the elevation of badge levels. Specifically, reviewers with higher badges put more effort in contributing reviews, such as posting more, longer, and easier to read reviews. 2) reviewers change their opinions as a consequence of badge upgrading. Concretely, reviewers with high level badges are more generous as they dislike posting extremely negative ratings or writing negative textual reviews.

This paper is expected to make contributions to several research areas. First, it enhances our understanding on how the hierarchical incentive mechanism affects user contribution in online communities from both users' efforts and opinions. While prior studies have confirmed the positive effect of hierarchical badges on users' efforts (Cavusoglu et al., 2021; Goes et al., 2016), our study demonstrates that the hierarchical incentive system is a double-edged sword regarding user contribution. On one hand, hierarchical badges motivate users to make more efforts in contributing reviews; on the other hand, the benefit comes at the cost of inducing users to post more positive opinions. Second, it brings new theoretical insights to the research strand of gamified incentive by revealing the psychological mechanisms of how user behavior changes in a hierarchical badges system. We draw on the goal-setting theory and status hierarchies theory to explain user contribution in online communities, thus expanding the theories from organizational behavior (Anderson et al., 2015; Locke et al., 1981) to the context of online UGC.

Our findings should be of interest for online communities that rely on user contributions. Retaining users and motivating them to contribute content is critically important for such platforms to make success, otherwise under-supply may eventually result in the downfall of the platforms. In this paper, we provide evidence for the effect of hierarchical incentive mechanism on user effort and opinion, which help the UGC platforms to design their incentive mechanisms. Particularly, employing a hierarchical badges system is helpful to increase users' efforts, including their review frequency, length, and readability. This echoes the popularity of hierarchical badges system in practice (Goes et al., 2016). However, our study also shows that a hierarchical badges system distorts users' opinions. Reviewers with higher badges tend to post fewer extremely negative ratings and use fewer negative words in their textual reviews, making their opinions more positive. This finding is particularly important since online review platforms are overwhelmingly positive and are suffering from the issue of authenticity (Hu et al., 2009; Schoenmueller et al., 2020). Such positive bias could negatively affect the platforms. Moreover, as Li et al. (2019) suggested, compared with reviewers with lower badges, reviewers with higher badges have greater impact on potential consumers as they are more likely and able to express their opinions, thus exacerbating the negative consequence of such positive bias. For incentive designers, they should be aware that the platform enjoys more user effort at the cost of biased user opinions. Motivating users to express their opinions truthfully would be a challenging topic. In addition, this study reveals a limitation of the hierarchical badges system, which is the negative effect of time on the effort of users. We found that the longer a user is on the site (registration time), the fewer and shorter reviews that he/she contributes. It appears that the more time passes, the less motivated users become. Exploring the reasons why users become less motivated as the time passes can be an interesting area for future research.

6. Limitations and future work

Our study is particularly important given that many online platforms are employing the hierarchical badges system as an incentive mechanism. The findings also draw insights for other contexts of UGC, such as social question and answer sites, social media, and crowdsourcing platforms. However, several limitations of this study are noteworthy and deserve further research. First, our study is based on a cross-sectional dataset, which does not allow us to examine the short effect and long effect on user behavior from a longitudinal view. Further study could examine an individual user's behavioral change in a hierarchical badges system from a fine-grained way by obtaining a panel dataset. Second, we use review frequency, length, and readability as the metrics of user effort. As suggested in prior studies, other indicators could be also used to measure user effort, such as review usefulness (Liu et al., 2016) and lexical density (Zhang et al., 2020). Third, we consider only the effects of the hierarchical badges system on users' contribution behavior, and not examine the downstream effect on users' consuming behavior. Studying the direct effect of such incentive mechanism on sales is an interesting avenue for further research as it is what the platform really cares about (Hamari, 2017). Fourth, understanding the effect of goal difficulty on user effort may also have important implications for platform practitioners (Gutt et al., 2020) but it is beyond the scope of the current study. As suggested by the goal-setting theory, a goal should not be too difficult or too easy so the participants would be more up for it. It would be interesting to examine the specific relationship between effort level and contribution points required to upgrade. Finally, some variables that may also affect user behavior were not included in our model, such as individual characteristics, e.g., gender, age, professional et al. as such information is private and always cannot be obtained on an online platform. It is worth combining multiple methods, such as surveys and field experiments to overcome the data limitations of online platforms.

Credit statement

Dandan Ma: Conceptualization, Methodology, Formal analysis, Writing - original draft. Shuqing Li: Data curation, Software. Jia Tina Du: Validation, Writing - original draft, Writing - review & editing. Zhan Bu: Methodology, Writing - review & editing. Jie Cao: Project administration, Funding acquisition. Jianjun Sun: Supervision, Resources.

Acknowledgements

This work was supported by the Key Program of National Natural Science Foundation of China (No: 92046026; No:91646204), the Philosophy and Social Science Foundation of the Jiangsu Higher Education Institutions (No: 2019SJA0272), and the Major Basic Research Project of the Natural Science Foundation of the Jiangsu Higher Education Institutions (No: 19KJA510011).

References

Anderson, C., & Brion, S. (2014). Perspectives on power in organizations. *Annual Review of Organizational Psychology and Organizational Behavior, 1*(1), 67–97.
Anderson, C., Hildreth, J. A. D., & Howland, L. (2015). Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological Bulletin, 141*(3), 574–601.
Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2013). Steering user behavior with badges[C]. *Proceedings of the 22nd international conference on world wide web*.
Antin, J., & Churchill, E. F. (2011). Badges in social media: A social psychological perspective. *CHI 2011 gamification workshop proceedings* (pp. 1–4).
Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117–148.
Bu, Z., Li, H. J., Zhang, C., Cao, J., Li, A., & Shi, Y. (2019). Graph k-means based on leader identification, dynamic game and opinion dynamics. *IEEE transactions on knowledge and data engineering*. http://doi.ieeecomputersociety.org/10. 1109/TKDE.2019.2903712.
Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (2018). Stimulating online reviews by combining financial incentives and social norms. *Management Science, 64*(5), 2065–2082.
Cabral, L., & Li, L. (2015). A dollar for your thoughts: Feedback-conditional rebates on eBay. *Management Science, 61*(9), 2052–2063.
Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems, 50*(2), 511–521.
Cavusoglu, H., Li, Z., & Kim, S. H. (2021). How do virtual badges incentivize voluntary contributions to online communities? *Information & Management, 58*(5), 103483.
Chen, W., Wei, X., & Zhu, K. (2018). Engaging voluntary contributions in online communities: A hidden markov model. *MIS Quarterly, 42*(1), 83–100.
Darejeh, A., & Salim, S. S. (2016). Gamification solutions to enhance software user engagement—a systematic review. *International Journal of Human-Computer Interaction, 32*(8), 613–642.
Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.
Dong, L., Huang, L., Hou, J. J., & Liu, Y. (2020). Continuous content contribution in virtual community: The role of status-standing on motivational mechanisms. *Decision Support Systems, 132*, 113283.
Drèze, X., & Nunes, J. C. (2011). Recurring goals and learning: The impact of successful reward attainment on purchase behavior. *Journal of Marketing Research, 48*(2), 268–281.
Fayazi, A., Lee, K., Caverlee, J., & Squicciarini, A. (2015). Uncovering crowdsourced manipulation of online reviews. *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval* (pp. 233–242).
Gallus, J. (2017). Fostering public good contributions with symbolic awards: A large-scale natural field experiment at wikipedia. *Management Science, 63*(12), 3999–4015.

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering, 23*(10), 1498–1512.
Goes, P. B., 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, P. B., Lin, M., & Au Yeung, C. M. (2014). "Popularity effect" in user-generated content: Evidence from online product reviews. *Information Systems Research, 25*(2), 222–238.
Gunning, R. (1969). The fog index after twenty years. *Journal of Business Communication, 6*(2), 3–13.
Gutt, D., von Rechenberg, T., & Kundisch, D. (2020). Goal achievement, subsequent user effort and the moderating role of goal difficulty. *Journal of Business Research, 106*, 277–287.
Hamari, J. (2017). Do badges increase user activity? A field experiment on the effects of gamification. *Computers in Human Behavior, 71*, 469–478.
Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? – a literature review of empirical studies on gamification. *Proceedings of the 47th Hawaii international conference on system sciences* (pp. 3025–3034).
Hanson, S., Jiang, L., & Dahl, D. (2019). Enhancing consumer engagement in an online brand community via user reputation signals: A multi-method analysis. *Journal of the Academy of Marketing Science, 47*(2), 349–367.
Huang, N., Burtch, G., Gu, B., Hong, Y., Liang, C., Wang, K., & Yang, B. (2018). Motivating user-generated content with performance feedback: Evidence from randomized field experiments. *Management Science, 65*(1), 327–345.
Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision Support Systems, 57*, 42–53.
Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM, 52*(10), 144–147.
Kemper, T. D. (1991). Predicting emotions from social relations. *Social Psychology Quarterly*, 330–342.
Keung, C., Lee, A., Lu, S., & O'Keefe, M. (2013). BunnyBolt: A mobile fitness app for youth. *Proceedings of the 12th international conference on interaction design and children*.
Khern-am-nuai, W., Kannan, K., & Ghasemkhani, H. (2018). Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research, 29*(4), 871–892.
Levina, N., & Arriaga, M. (2014). Distinction and status production on user-generated content platforms: Using Bourdieu's theory of cultural production to understand social dynamics in online fields. *Information Systems Research, 25*(3), 468–488.
Li, W., Grossman, T., & Fitzmaurice, G. (2012). GamiCAD: A gamified tutorial system for first time AutoCAD users. *25th annual ACM symposium on user interface software and technology*.
Li, Y., Kuan, K. K., & Liu, N. (2019). Exploring the linguistic characteristics of online consumer reviews by top reviewers and ordinary reviewers. *23th pacific asia conference on information systems*.
Liu, D., Santhanam, R., & Webster, J. (2017). Toward meaningful engagement: A framework for design and research of gamified information systems. *MIS Quarterly, 41*(4), 1011–1034.
Liu, X., Schuckert, M., & Law, R. (2016). Online incentive hierarchies, review extremity, and review quality: Empirical evidence from the hotel sector. *Journal of Travel & Tourism Marketing, 33*(3), 279–292.
Liu, X., Schuckert, M., & Law, R. (2018). Utilitarianism and knowledge growth during status seeking: Evidence from text mining of online reviews. *Tourism Management, 66*, 38–46.
Locke, E. A. (1967). Motivational effects of knowledge of results: Knowledge or goal setting? *Journal of Applied Psychology, 51*(4), 324–329.
Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance, 3*(2), 157–189.
Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin, 90*(1), 125–152.
Lu, S., Xie, Y., & Chen, X. (2019). *Dynamic effect of digital badges in user-generated content platforms*. https://ssrn.com/abstract=3560378.
Magee, J. C., & Galinsky, A. D. (2008). Social hierarchy: The self-reinforcing nature of power and status. *The Academy of Management Annals, 2*(1), 351–398.
Ma, D. D., Li, S. Q., Du, J. T., Lu, C., & Sun, J. J. (2020). Does hierarchical badges system affect users' efforts and opinions? A preliminary study. ASIST 2020. In *Proceedings of the 83rd annual meeting of the association for information science and technology* (Vol. 57, p. e345). https://doi.org/10.1002/pra2.345, 1.
Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science, 341*(6146), 647–651.
Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. Com. *MIS Quarterly*, 185–200.
Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research, 50*, 67–83.
Perretti, F., & Negro, G. (2006). Filling empty seats: How status and organizational hierarchies affect exploration versus exploitation in team design. *Academy of Management Journal, 49*(4), 759–777.
Qiao, D., Lee, S. Y., Whinston, A. B., & Wei, Q. (2020). Financial incentives dampen altruism in online prosocial contributions: A study of online reviews. *Information Systems Research, 31*(4), 1361–1375.
Schoenmueller, V., Netzer, O., & Stahl, F. (2020). The polarity of online reviews: Prevalence, drivers and implications. *Journal of Marketing Research, 57*(5), 853–877.
Staw, B. M. (1976). Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance, 16*(1), 27–44.
Stephen, A., Bart, Y., Du Plessis, C., Goncalves, D., Gürhan-Canli, Z., Otnes, C., & Zhu, R. (2012). Does paying for online product reviews pay off? The effects of monetary incentives on content creators and consumers. In *, Vol. 40. Advances in consumer research* (pp. 228–231). Duluth: MN: Association for Consumer Research.
Su, C., & Cheng, C. H. (2015). A mobile gamification learning system for improving the learning motivation and achievements. *Journal of Computer Assisted Learning, 31*(3), 268–286.
Sun, Y., Dong, X., & McIntyre, S. (2017). Motivation of user-generated content: Social connectedness moderates the effects of monetary rewards. *Marketing Science, 36*(3), 329–337.
Tondello, G. F., Premsukh, H., & Nacke, L. E. (2018). A theory of gamification principles through goal-setting theory. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 1–14.
Tseng, C. W., Chou, J. J., & Tsai, Y. C. (2018). Text mining analysis of teaching evaluation questionnaires for the selection of outstanding teaching faculty members. *IEEE Access, 6*, 72870–72879.
Wang, Y., Lu, X., & Tan, Y. (2018). Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines. *Electronic Commerce Research and Applications, 29*, 1–11.
Wang, D., Lee, Y. C., & Fu, W. T. (2019). I love the feeling of being on stage, but I become greedy" exploring the impact of monetary incentives on live streamers' social interactions and streaming content. *Proceedings of the ACM on Human-Computer Interaction*, 1–24.
Willer, R. (2009). Groups reward individual sacrifice: The status solution to the collective action problem. *American Sociological Review, 74*(1), 23–43.
Zhang, M., Wei, X., & Zeng, D. D. (2020). A matter of reevaluation: Incentivizing users to contribute reviews in online platforms. *Decision Support Systems, 128*, 113–158.