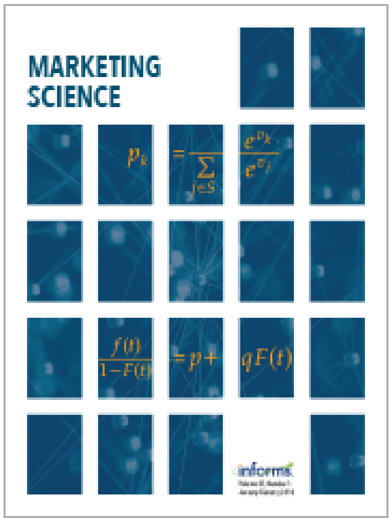


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Do Larger Audiences Generate Greater Revenues Under Pay What You Want? Evidence from a Live Streaming Platform

Shijie Lu , Dai Yao , Xingyu Chen , Rajdeep Grewal

informs
http://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Vol. 40, No. 5, September–October 2021, pp. 964–984
ISSN 0732-2399 (print), ISSN 1526-548X (online)

Do Larger Audiences Generate Greater Revenues Under Pay What You Want? Evidence from a Live Streaming Platform

Shijie Lu,[a] Dai Yao,[b] Xingyu Chen,[c,*] Rajdeep Grewal[d]

[a] University of Houston, Houston, Texas 77004; [b] National University of Singapore, Singapore 119245; [c] Shenzhen University, Shenzhen, Guangdong Province 518060, China; [d] University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599
*Corresponding author

Contact: slu@bauer.uh.edu, https://orcid.org/0000-0002-4180-6022 (SL); dai.yao@nus.edu.sg, https://orcid.org/0000-0002-0852-0700 (DY); celine@szu.edu.cn, https://orcid.org/0000-0002-3813-0680 (XC); grewalr@unc.edu, https://orcid.org/0000-0003-4467-2717 (RG)

Received: January 6, 2020
Revised: July 21, 2020; November 16, 2020
Accepted: December 27, 2020
Published Online in Articles in Advance: August 19, 2021

https://doi.org/10.1287/mksc.2021.1292

Abstract. As live streaming of events gains traction, pay what you want (PWYW) pricing strategies are emerging as critical monetization tools. We assess the viability of PWYW by examining the relationship between popularity (i.e., audience size) of a live streaming event and the revenue it generates under a PWYW scheme. On the one hand, increasing audience size may enhance voluntary payment/tips if social image concerns are important because larger audiences amplify the utility pertaining to social image. On the other hand, increasing audience size may reduce tips if gaining the broadcaster's reciprocal acts motivates tipping because larger audiences are associated with fiercer competition for reciprocity. To examine these trade-offs in the relationship between audience size and revenue under PWYW, we manipulate audience size by exogenously adding synthetic viewers in live streaming shows on a platform in China. The results reveal a mostly positive relationship between audience size and average tip per viewer, which suggests that social image concerns dominate seeking reciprocity. In support of herding, adding synthetic viewers also increases the number of real viewers. Social image concerns and herding together explain the finding that adding one additional viewer improves the tipping revenue per minute by approximately 0.01 yuan (1% of the mean level). Further, famous female broadcasters who use recognition-related words frequently during the event benefit the most from an increase in audience size. Overall, the results indicate that revenues under PWYW do not scale linearly and support the relevance of social image concerns in driving individual payment decisions under PWYW.

History: K. Sudhir served as the senior editor and Juanjuan Zhang served as associate editor for this article.
Funding: This research is supported by the National Natural Science Foundation of China [Grant 71872115], the Natural Science Foundation of Guangdong Province, China [Grant 2020A1515011201], and National University of Singpore Research [Grant R-316-000-104-133].
Supplemental Material: The data and online appendix are available at https://doi.org/10.1287/mksc.2021.1292.

Keywords: live streaming • pay what you want • social media • user-generated content • tipping • field experiment • video analysis

1. Introduction

Peer-to-peer live streaming (live streaming hereinafter) of events, such as sports (e.g., video games; The Economist 2014), hobbies (e.g., bird watching; Knowledge@Wharton 2015), and political commentaries (e.g., satirical videos; Qin 2016), is the next big revolution in democratizing the production and broadcast of videos. Not surprisingly, live streaming is getting attention from major players: the market has witnessed the launch of Facebook Live and YouTube Live Streaming Channels, the acquisition of Twitch by Amazon.com, and the emergence of Twitter's Periscope. According to Facebook, these live videos are 10 times more likely to generate comments than recorded videos and viewed three times longer than regular videos (Mediakix 2018). The growth of live streaming in China is phenomenal and outstrips other countries (D'Urbino 2017); the celebrity of live streaming stars parallels those of movie stars (China Daily 2016). The size of the live streaming market in China reached $4.4 billion in 2018, a 32% increase over 2017, and the size of the global live streaming market was $7.4 billion in 2018, a 47% increase over the previous year (Deloitte 2018).

Two monetizing strategies are popular with user-generated content (UGC), which includes live streaming. The indirect model involves advertising and product placements, and the direct model involves charging the viewers. The nascent stages of live streaming and the transient nature of live streaming viewers have resulted in firms testing a pay what you want (PWYW) pricing strategy (e.g., Kim et al. 2009, Gneezy et al. 2012). Under PWYW, viewers can access live video content free of charge and pay/tip broadcasters in real time by sending voluntary payments in the form of virtual cash and virtual gifts. The revenue collected from viewers is later split between the live streaming platform and broadcasters. Prominent live streaming platforms, such as YouTube Live and Twitch, fully embrace the PWYW strategy with Facebook Live adopting this pricing strategy for live streaming of video games (Roettgers 2018).

The prominence of the PWYW revenue model leads to astronomical growth of the live streaming industry in both the United States and China over the past few years. According to Streamlabs and Goldman Sachs research, live streaming broadcasters within the United States received $129 million in the form of discretionary tips in 2017 (Hays 2018). It also predicts that the "tipping market" will reach $372 million in 2022, suggesting a compound annual growth rate of 23.6% from 2017 to 2022. For the live streaming market in China, the user base has increased from 310 million in 2016 to 504 million in 2019, suggesting that approximately two out of five Chinese watched live streaming shows in 2019 (iiMedia Research 2020). With the booming live streaming industry, a number of Chinese live streaming service providers (e.g., Bilibili, DouYu, Huya, Momo) have become billion-dollar companies that are publicly traded in the United States.

Despite the wide adoption of PWYW in the live streaming industry, the viability and efficacy of this revenue model remain unclear. In this research, we assess the scalability of PWYW by addressing the following question: do larger audiences generate greater revenues in live streaming with PWYW schemes? As tipping revenue in a live streaming session depends on the number of viewers and tip amount per viewer, we first ask how popularity information (i.e., audience size) affects viewer participation. In settings in which product quality is unknown to consumers a priori, popularity can serve as a signal of quality to draw in new viewers (e.g., Tucker and Zhang 2011, Zhang and Liu 2012). Because of the self-reinforcement of popularity information, we expect herding to occur in the context of live streaming, suggesting a positive effect of audience size on viewer participation.

Further, we examine whether larger audiences enhance or reduce the average tip amount per viewer. We posit that the direction of this effect depends on a central trade-off. On the one hand, theory of signaling and status seeking (e.g., Lampel and Bhalla 2007, Gneezy et al. 2012, Toubia and Stephen 2013) and the public nature of viewers' payment in live streaming suggest that a viewer's utility of tipping should increase as audience size increases because of an upward bump in the viewer's *social image*, defined as an individual's social status and prestige perceived by others (Lampel and Bhalla 2007). As such, a representative viewer's tip amount should increase as audience size grows. On the other hand, if *seeking reciprocity* is the primary motive for tipping, as a session becomes more crowded, the chance of a viewer gaining reciprocal acts from the broadcaster in the form of social interactions should reduce, which should lead to a lower tip per viewer because of the more intense (perceived) competition for reciprocity (referred to as the *N-effect* in social psychology; Garcia and Tor 2009). The negative relationship between audience size and tip amount per viewer might also occur because of free-riding: as group size grows, the individual contribution declines (e.g., Olson 1965, Andreoni 1988). Thus, the net effect of audience size on tip amount per viewer is a priori unclear. It could be either positive, negative, or null, depending on the relative strengths of multiple underlying forces (e.g., social image, reciprocity, free-riding).

Two features of the live streaming industry make it suitable for us to study the scalability of the PWYW revenue model. First, both the payment decisions and audience size are public information, which enable status-signaling and herding by viewers. These behavioral mechanisms are likely absent for products and services that do not disclose consumers' payment to others (e.g., shows on Netflix). Second, a live streaming event is hosted over the internet and, therefore, does not have a capacity constraint on the number of consumers who can simultaneously use the service. This is different from off-line events (e.g., sports games, concerts) for which the number of consumers is restricted by the physical space.

To understand how audience size affects revenues under PWYW, we conduct a field experiment on a large live streaming platform in China. In the experiment, we first randomize the displayed audience size (based on treatment and control condition allocation) across broadcasters. Then, using the allocation of broadcasters to one of the three conditions (two treatment and one control), we randomize the audience size within broadcaster, within session (as most broadcasters have multiple sessions of around an hour each), and at every minute of a session (after the first 10 minutes as we discuss subsequently), for which deviation of displayed audience size from the actual audience size depends on a random draw. Specifically, for each of the two treatment conditions, at every minute, we draw a random number of synthetic viewers to add from a distribution with a mean of two or four, respectively.

Empirical identification of the treatment effect of adding synthetic viewers proceeds in three steps. First, we begin with a mean comparison by exploiting variation at the broadcaster level across treatment groups. The results from the mean comparison suggest that adding synthetic viewers indeed significantly improves tipping revenues per minute; however, such a benefit from audience augmentation is subject to diminishing returns as the difference is not statistically significant when increasing the average number of synthetic viewers per minute from two to four.

Second, though intuitive and convenient, the mean comparison at the broadcaster level is subject to aggregation bias mainly because the treatment strength varies across sessions and within sessions over time; thus, averaging could result in aggregation bias. To tackle this problem, we resort to the slope comparison by testing whether the slope of tipping revenues against time differs across treatment conditions because the average treatment strength increases over time from our manipulation (i.e., we add an average of two or four viewers every minute based on treatment condition). The slope analysis yields qualitatively similar results with increased statistical significance of identified effects.

Third, the panel structure of the data allows us to precisely account for the time-varying treatment strength randomized at the minute level to obtain an unbiased estimate of the treatment effect. We also control for unobserved session heterogeneity with session fixed effects (which alleviates the need for broadcaster-level fixed effects) to identify the effect of audience size based on within-session variation across time. Several findings emerge from the linear panel regressions. Increasing the audience size by one unit improves the tipping revenue per minute by approximately 0.01 yuan, which is 1% of the mean level. Such a positive effect on tipping revenue is subject to diminishing returns, and the effect turns negative if the number of displayed viewers exceeds 567 (96.6 percentile). By breaking down the revenue to the number of real viewers times tip amount per viewer, we find that two forces drive the mostly positive relationship between average treatment strength and tipping revenue: a positive treatment effect on viewer size and a mostly positive treatment effect on tip per viewer. The former effect on viewer size seems to rely on theories related to herding, and the latter effect on tip per viewer suggests the dominance of social image concerns over seeking reciprocity in driving tipping decisions.

We explore the potential heterogeneity in the effect of increasing audience size in subsequent analyses. We find that female and more famous broadcasters tend to enjoy greater benefits from larger audiences than male and less famous broadcasters. Such an improvement in tipping revenue comes from both a stronger effect on audience size and tip per viewer. These results are generally consistent with the data pattern predicted by herding resulting from social norm (Croson and Shang 2008, Simonsohn and Ariely 2008) rather than observational learning (Banerjee 1992, Bikhchandani et al. 1992). Furthermore, the theory of intrasexual competition from evolutionary psychology predicts that status-signaling motivation among male customers is more likely to appear in the presence of a female rather than a male broadcaster (Sundie et al. 2011). Given the majority of male audiences on our platform (66%), our finding of a stronger effect on tipping to female than male broadcasters suggests the relevance of social image concerns in live streaming. We further provide corroborating evidence of social image–related utility in tipping by examining the moderating effects of broadcaster performances. If social image concerns are indeed important, we expect the broadcaster's tendency to express recognition of individual viewers to strengthen the positive effect of audience size on tipping. We employ state-of-the-art speech recognition techniques to measure the broadcaster's use of recognition-related words during a live streaming event and find the positive moderating effect predicted by the model with social image–related utility.

With this research, we aim to make several important contributions. Substantively, we find an overall positive and concave causal relationship between audience size and tipping revenue, suggesting that the revenue under PWYW is less scalable than that under other monetization tools (e.g., advertising, product placement) that sell viewers' eyeballs or clicks, which increase linearly with the audience size. To our knowledge, this research is among the first few field data–based as opposed to survey-based studies in the live streaming literature. With our finding of a positive effect of audience size on average tip amount per viewer, this research also contributes to the PWYW literature by confirming the relevance of social image–related utility in commercial contexts. Although research shows the relevance of social image motive in charity contexts (Harbaugh 1998, Ariely et al. 2009, DellaVigna et al. 2012), no evidence exists of social image concerns or social pressure in PWYW employed in commercial contexts. Our data from live streaming provide new evidence of the importance of social image concerns in driving individual payment decisions under PWYW.

2. Related Literature

As live streaming is an emerging form of UGC, the literature on UGC is of relevance. Previous research shows the importance of social influence in driving individual contribution on various types of UGC platforms. For example, Toubia and Stephen (2013) conduct a field experiment by adding synthetic followers on Twitter and find that social image–related utility

(utility associated with the perception of others) is generally more important than intrinsic utility (direct utility from posting content) in motivating users to contribute content; thus, users' content contribution is likely to increase with the augmentation of followers. Zhang and Zhu (2011) find social benefit in content generation by examining users' posting and editing efforts on Wikipedia. Shriver et al. (2013) advance the literature by showing the reverse causal effect of content generation on user engagement and social ties in the context of an online windsurfing community. Live streaming differs from conventional UGC in its core business model. Almost all UGC platforms studied (e.g., online forums, review sites, social network) monetize online traffic through advertisements. By contrast, most live streaming platforms rely on viewers' voluntary payments/tips to generate revenues. Our study differs from previous research on UGC by exploring the scalability of revenues against the size of consumers when the UGC platform monetizes traffic through PWYW.

As voluntary payments/tips facilitate the pricing of live streaming, our study closely relates to the literature on tipping and PWYW. Tipping is a common and important component of service marketing (for a review, see Azar 2007); for example, annual tipping in the U.S. food industry amounts to $46.6 billion (Azar 2011). There is increasing interest from both industry and academia in exploring whether tipping, in lieu of a fixed price, can serve as an alternative business model to generate revenues (e.g., Natter and Kaufmann 2015). The English rock band Radiohead is a famous application of this idea. Radiohead announced in 2007 that it would allow fans to set their own price, if anything, for downloading its seventh album *In Rainbows*. The live streaming industry is also adopting and experimenting with the viability of this PWYW business model.

Such PWYW pricing schemes, in which consumers decide the payment amount (including zero), are beginning to receive scholarly scrutiny (e.g., Schmidt et al. 2015, Jung et al. 2016, Chen et al. 2017). Kim et al. (2009) find that PWYW transactions evoke concerns about reciprocity and fairness. Gneezy et al. (2012) identify the importance of self-identity and self-image in PWYW and show that self-image concerns drive people to pay more but also make them less likely to buy. Research also shows that the strategy of shared social responsibility, a modified version of PWYW with a fixed percentage of consumers' payment going to support a charity cause, is more profitable than a fixed price (Gneezy et al. 2010) and insensitive to the percentage of payment allocated to the charity (Jung et al. 2017). Our study differs from previous studies of PWYW in that the tip by consumers is publicly observable in live streaming instead of anonymous as in off-line settings (e.g., theme parks, restaurants) in which previous PWYW experiments typically occurred. The public nature of payment in PWYW in the live streaming industry suggests the importance of social image, which drives an individual's desire to improve the social status and prestige perceived by others (Lampel and Bhalla 2007). We contribute to the PWYW research by providing suggestive evidence of the existence of social image–related utility in driving consumers' payment in the noncharity contexts. This novel finding of social image–related utility under PWYW in the live streaming industry is by and large consistent with the prominence of social image in motivating user contributions in front of others on online platforms (Toubia and Stephen 2013).

The total revenue in live streaming depends not only on the individual tip amount but also on the size of viewers. Thus, we draw from previous herding literature to theorize the impact of popularity information on viewer participation (e.g., Banerjee 1992, Bikhchandani et al. 1992). Before entering a live stream, viewers have only limited information about its quality and, thus, may resort to other participants to rationalize their decisions, resulting in herding behavior, which predicts a positive effect of popularity information on viewer participation. Previous literature further suggests that herding occurs for at least two reasons: (1) social norm, in which a potential viewer refers to existing viewers' choices as a descriptive social norm and then passively mimics their decisions (e.g., Croson and Shang 2008, Simonsohn and Ariely 2008), and (2) observational learning, in which a potential viewer infers higher quality from a session with a larger number of existing viewers and, thus, increases the likelihood of participation (e.g., Cai et al. 2009, Zhang 2010). In this study, we find data patterns that are more consistent with herding driven by social norm rather than observational learning.

Finally, our research builds on the growing literature on live streaming. Extant research is largely computer science focused (e.g., Pires and Simon 2015, He et al. 2016) with only a handful of studies examining the reasons behind the production and popularity of live streaming. With a survey of Twitch viewers, Sjöblom and Hamari (2017) identify tension release, interpersonal bonding, and entertainment as three key motivators to watch live streaming of video games. Lee et al. (2019) survey live streaming viewers in China and find that interaction and content appear to be the two most important reasons that motivate viewers to tip. For interactions, they further find that viewers are motivated by both reciprocity, that is, viewers expect broadcasters to engage in social interactions with tippers in return for tips received, and social image, that is, viewers tip to grab attention from the crowd and enjoy standing out from the others.

Another survey from an industrial report of Chinese live streaming platforms also highlights the importance of social image concerns in viewers' tipping (iiMedia Research 2019). According to this survey, receiving social recognition is among the top five reasons why viewers tip in live streaming. The other four reasons include content uniqueness, content relevance, broadcaster charisma, and social norm. For broadcasters' behaviors, Tang et al. (2016) interview 20 frequent broadcasters on Meerkat and Periscope, two popular live streaming apps in the United States, and find that most broadcasters use live streaming to build their personal brand. In contrast with previous survey-based studies, we execute a randomized field experiment rather than self-reports to understand tipping behaviors manifested in individuals' real actions.[1]

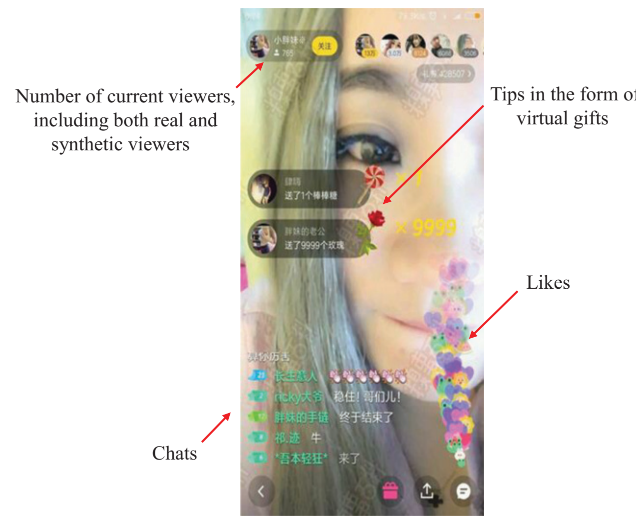
3. Field Setting and Experiment Design

3.1. Background

We conduct our study in collaboration with one popular peer-to-peer live streaming platform in China, which prefers to remain anonymous. This platform started in 2005 as an online community in which users could post jokes and share entertaining stories. As internet users substantially shifted their time from desktop to mobile devices, the platform launched a mobile app in 2012 that provided a service similar to its website. In early 2016, the platform expanded its product line by offering a new live streaming service in its existing mobile app. By the time of our study in August 2016, this live streaming app had more than 600,000 monthly active viewers and more than 40,000 registered broadcasters. In general, the broadcasters of this live streaming app are young with 55% aged 18–24, 40% aged 25–34, and 5% aged 35–44 years. Approximately 80% of the broadcasters are female, and 66% of viewers are male.[2]

The live streaming app used in our study is typical of its counterparts in the United States. When opening the app, a viewer can see thumbnails of live sessions and the number of people currently watching each session. At the time of our study, recency of starting time determined session ordering. Viewers are free to join any session hosted by a broadcaster. During each session, viewers can interact with the broadcaster and the audience in three ways: using *tips*, *chats*, and *likes*. Figure 1 shows an example of a viewer interface during a live streaming session. Here, viewers can send tips in the form of virtual gifts purchased through the app. These virtual gifts appear on the screen for about one to five seconds, depending on the gift value. The prices of these gifts range from 0.1 to 1,000 yuan.[3] The platform pays a fixed proportion of the revenue from those virtual gifts to the broadcaster. The types of broadcasting content are reality shows of broadcasters, who

Figure 1. (Color online) A Snapshot of a Live Streaming Session



usually chat with the audience about trendy topics and occasionally perform (e.g., singing, playing music). Online Appendix A provides details on the broadcasting genres. An important feature of this live streaming app is its real-time updates to the number of viewers in each session, and this information is public. There is no capacity constraint on the maximum number of viewers in a session, which allows us to conduct experiments by manipulating the audience size.

3.2. Experiment Design

We designed and implemented a randomized controlled field experiment to estimate the causal effect of audience size on tipping revenues. We collaborated with the focal live streaming app to add exogenous variation in the number of viewers during broadcasting sessions. Before our experiment, the platform provided us with live streaming data from a random sample of 165 active broadcasters during the period from July 8 to August 8, 2016. For each session, we observed the session length; the number of tips, chats, and likes; and the number of chatters, which represents the number of viewers who submitted chats at least once during a session.[4] We randomly assigned those broadcasters into three groups. Although, initially, each group had 55 broadcasters, a few of them did not broadcast during our experiment from August 11 to September 12, 2016. This lack of broadcasting left the first group with 48 broadcasters, the second group with 51, and the third group with 54. We set live streaming sessions by the first group as the control group and treated those by the second and third groups, which we call treatment groups 1 and 2 (T1 and T2) hereinafter.

To check whether our randomization worked as intended, we compared the mean of key pretreatment metrics at the level of broadcasters across the three

Table 1. Means of Pretreatment Variables Across Groups

	Means			
	C	T1	T2	p-values of F-test
Tips per session (yuan)	24.7	26.5	24.4	0.973
Number of chats per session	193.2	166.9	183.7	0.756
Number of likes per session	3,723.8	4,409.2	4,275.6	0.931
Number of chatters per session	21.7	19.7	22.4	0.883
Session length (minutes)	51.0	46.3	46.4	0.622
Number of sessions	16.5	13.3	14.0	0.274
Female	0.771	0.863	0.815	0.502
Number of broadcasters	48	51	54	

groups. These metrics included the average value of tips, the average number of likes and chats, and the average number of chatters per session. We also compared the number of sessions, session length, and proportion of female broadcasters across the three groups. As Table 1 shows, the p-values of the F-test indicate no significant evidence to reject the null hypothesis that the means of the pretreatment metrics by broadcasters in each group are the same, thus confirming the success of our randomization procedure.

For each session by a broadcaster in T1, we asked the platform to add an average of two synthetic viewers at the end of each minute after the 10th minute. We treat sessions in T2 similarly to those in T1 but double the strength of the treatment by adding an average of four synthetic reviews at each minute after the 10th minute. The platform added $\lfloor x \rfloor$ synthetic viewers at the end of each minute; x is drawn from a normal distribution, and $\lfloor \cdot \rfloor$ is the floor function. When $\lfloor x \rfloor$ is negative, up to $\lfloor x \rfloor$ number of synthetic viewers leave the session. We did not add synthetic viewers at the beginning of a session to avoid potential suspicion from the broadcaster and real viewers. The platform also used only dormant accounts—those generated by real users who were no longer active—as synthetic viewers in this study. Thus, the platform did not create any new accounts for this study, which could be troublesome if a savvy viewer noticed any unauthentic profile information such as a relatively recent registration date. The synthetic viewers we added did not interact with any party during a session. Figure 2 shows the distribution of the number of added synthetic viewers at each minute in T1 ($M = 2.08$) and T2 ($M = 4.12$), respectively. This manipulation allows us to have exogenous data variation at the level of minutes though we conduct initial randomization at the level of broadcasters.

Recall that the goal of our research is to investigate the relationship between audience size and tipping revenue, conditional on everything else, including engagement activities (i.e., tips, chat, likes). We, therefore, did not allow synthetic viewers to engage in a live stream so that we can have a relatively clean setting to identify the effect of audience size alone rather than a combined effect of audience size and engagement activities. It is also technically challenging to generate synthetic viewers who can meaningfully send tips, chats, and likes as a regular viewer. Thus, we manipulated only the audience size but not any engagement activity in this experiment. This design is similar to Toubia and Stephen (2013) in which the authors manipulated only the number of followers rather than their activities (e.g., replying, sharing, mentioning) on Twitter when studying the effect of followers on content creation.

3.3. Data Description

Our data include 2,222 sessions by 153 broadcasters during the experiment and 2,226 sessions by these broadcasters before the experiment.[5] We have 813, 660, and 749 sessions in C, T1, and T2, respectively, during the experiment. The number of sessions in C, T1, and T2 are 794, 678, and 754, respectively, before

Figure 2. (Color online) Histogram of Number of Added Synthetic Viewers

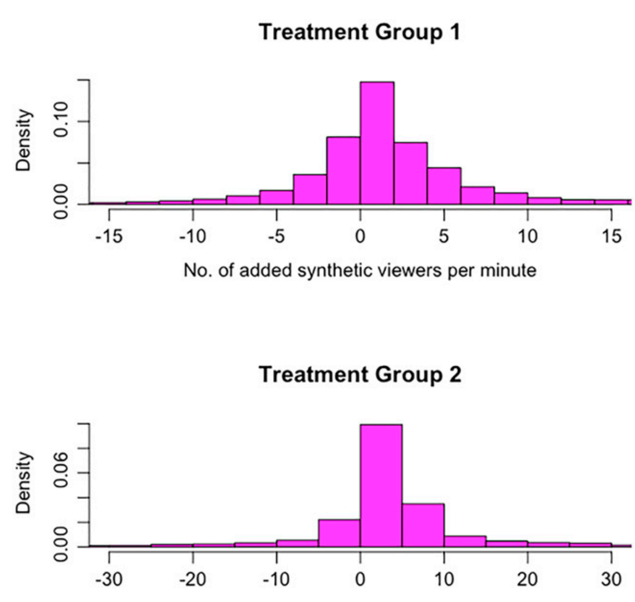


Table 2. Summary Statistics of Viewer-Related Variables Across Groups

	Definition	Group	Mean	Standard deviation	Minimum	Maximum
Tip_{kt}	Total tips received at minute t of session k, cents	C	61.2	480	0	30,000
		T1	103	609	0	16,810
		T2	115	822	0	30,000
$TipRate_{kt}$	Average tips per real viewer at minute t of session k, cents	C	5.20	29.3	0	1,680
		T1	6.92	33.8	0	3,011
		T2	5.86	29.8	0	3,033
$ChatRate_{kt}$	Average number of chats per real viewer at minute t of session k	C	0.572	0.468	0	21.8
		T1	0.515	0.410	0	25.5
		T2	0.529	0.508	0	35.5
$LikeRate_{kt}$	Average number of likes per real viewer at minute t of session k	C	4.36	7.54	0	290
		T1	4.86	7.94	0	254
		T2	3.24	6.39	0	204
N_{kt}^{real}	Number of real viewers at minute t of session k	C	9.96	10.1	1	200
		T1	11.9	12.2	1	207
		T2	15.3	18.6	1	492
N_{kt}^{syn}	Number of synthetic viewers at minute t of session k	C	0	0	0	0
		T1	72.4	67.7	0	694
		T2	139	137	0	2,558
N_{kt}^{disp}	Number of displayed viewers at minute t of session k	C	9.96	10.1	1	200
		T1	84.3	70.9	0	710
		T2	155	142	1	2,558

Notes. $TipRate = \frac{Tip}{N^{real}}$, $N^{disp} = N^{real} + N^{syn}$. 1 cent = 0.01 yuan.

the experiment. In contrast with session-level data collected before the experiment, we have minute-level observations during the experiment. In particular, we observe the number of likes and chats and the value of tips sent by real viewers during each minute t of a session k. We also observe the number of real and synthetic viewers at minute t. For broadcasters, we observe session length and infer broadcaster gender on the basis of saved video recordings of all streams. In Section 5.2, we describe how we generate four performance-related metrics by analyzing video content (i.e., *Recognition*, *FaceTime*, *HandTime*, and *Emotion*). We present the summary statistics of key variables for viewers and broadcasters in Tables 2 and 3, respectively. On average, each session lasts approximately 50 minutes, and a broadcaster in C, T1, and T2 receives 0.61, 1.03, and 1.15 yuan of tips per minute, respectively.

Table 3. Summary Statistics of Broadcaster-Related Variables Across Groups

	Definition	Group	Mean	Standard deviation	Minimum	Maximum
$Recognition_{kt}$	Count of four recognition-related words: "welcome," "hello," "thank you," and "many thanks" during a minute	C	6.09	5.98	0	49
		T1	5.60	5.73	0	48
		T2	6.03	6.23	0	47
$FaceTime_{kt}$	Duration of the broadcaster's face appearing in front of the camera during a minute	C	0.644	0.376	0	1
		T1	0.805	0.302	0	1
		T2	0.720	0.347	0	1
$HandTime_{kt}$	Duration of the broadcaster's hands appearing in front of the camera during a minute	C	0.281	0.311	0	1
		T1	0.415	0.355	0	1
		T2	0.324	0.319	0	1
$Emotion_{kt}$	Emotion score of the broadcaster during a minute: –1 means pure negativity, and 1 means pure positivity	C	0.095	0.183	–1	1
		T1	0.102	0.150	–0.983	1
		T2	0.121	0.185	–0.933	1
$Length_k$	Length of session k, minutes	C	50.3	52.4	1	345
		T1	52.6	53.0	1	363
		T2	54.6	59.1	1	687
$Female_i$	Gender of broadcaster i (1 = female, 0 = male)	C	0.771	0.425	0	1
		T1	0.862	0.238	0	1
		T2	0.815	0.392	0	1

Table 4. Correlation Matrix

	Tip	*TipRate*	*ChatRate*	*LikeRate*	N^{real}	N^{syn}	N^{disp}
Tip	1						
TipRate	0.733**	1					
ChatRate	−0.054**	−0.007*	1				
LikeRate	−0.000	0.023**	0.111**	1			
N^{real}	0.171**	0.040**	−0.259**	−0.055**	1		
N^{syn}	0.037**	0.011**	−0.007*	−0.011**	0.198**	1	
N^{disp}	0.059**	0.016**	−0.042**	−0.019**	0.326**	0.991**	1

**$p < 0.01$; *$p < 0.05$.

Table 4 reports the correlation matrix of key variables in our data. We find a positive correlation between the tip amount and the audience size measured by the number of displayed viewers ($r = 0.059$, $p < 0.01$). This positive association seems to be driven by the positive correlation between audience size and the number of real viewers ($r = 0.326$, $p < 0.01$) and the positive correlation between audience size and the average tip per viewer measured by the *TipRate* ($r = 0.016$, $p < 0.01$).

We show how the distributions of the three revenue-related variables (i.e., *Tip*, *TipRate*, and N^{real}) vary by the number of synthetic viewers in Figure 3. We use a heat map to visualize the distribution; the darker areas represent the higher density. The scatterplots of our raw data reveal an inverted U-shaped associations between N^{syn} and *Tip* and between N^{syn} and *TipRate* and a mostly positive and concave relationship between N^{real} and N^{syn} when the number of synthetic viewers is not too large ($N^{syn} \leq 1,200$). Figure 3 also depicts outliers in which the number of added synthetic viewers is extremely large ($N^{syn} > 1,200$) because of a few unusually long sessions.

4. Main Analysis

4.1. Mean Comparison

We describe the raw treatment effect by comparing the means of revenue-related variables across C, T1, and T2. Specifically, we examine whether the value of tips per minute, the session length, and the streaming frequency, all aggregated at the broadcaster level, vary by treatment conditions. We decompose the value of tips to the number of real viewers (N^{real}) and average tip per real viewer (*TipRate*) to explore the potential mechanism through which audience size may affect tipping revenue. As previously theorized, the popularity information may affect tipping revenues through the impact on viewer participation because of herding and through the impact on individual tipping because of the potential coexistence of viewers' motivations of improving social image and seeking reciprocity.

We report the results of the broadcaster-level mean comparison in Table 5. We note that the value of tips per minute (i.e., *Tip*) increases significantly when we add synthetic viewers. However, the value of tips is not significantly different between the T1 and T2 conditions in which treatment strength doubles from T1 to T2. This result suggests that the positive effect of treatment on *Tip* is nonlinear and subject to diminishing returns. By decomposing *Tip* into N^{real} times *TipRate*, we show that two factors drive the positive treatment effect on tipping revenue: a positive treatment effect on viewer participation and a positive treatment effect on *TipRate*. With regard to the length and frequency of streams, we find no significant difference between the control and treated broadcasters. We speculate that the null effect of audience size manipulation on session length might depend on broadcasters' schedules. Most broadcasters on our live streaming platform are using the streaming services either as a hobby or a part-time job. They typically have other duties (e.g., full-time worker/student) that may affect their flexibility of broadcasting. We also observe that broadcasters often announce how much time they will broadcast at the beginning of a session. The lack of flexibility and predetermined broadcasting schedules might explain the similar session length between the control and treated conditions.

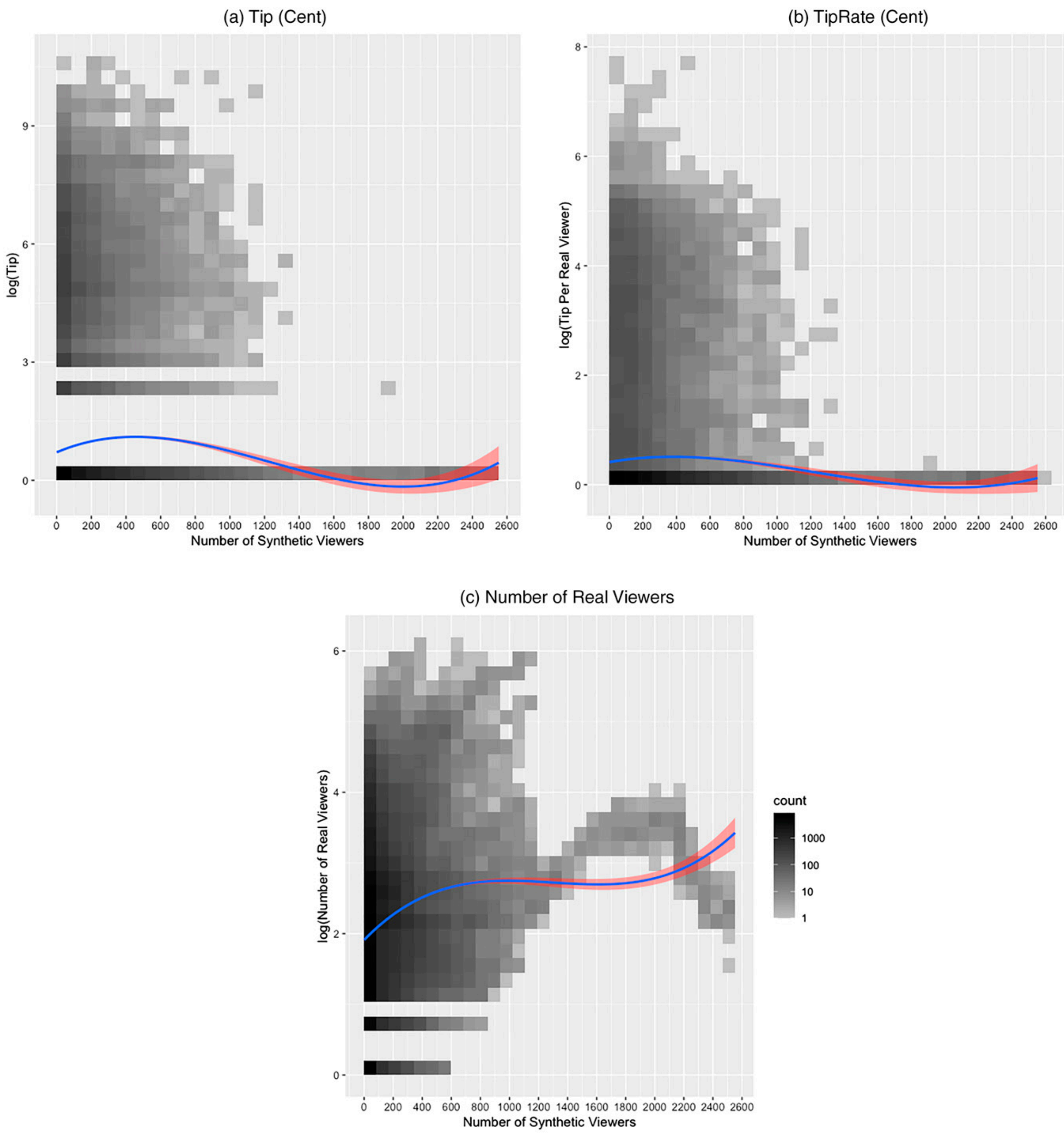
Despite the convenience, the treatment effect revealed from the aggregate-level mean comparison suffers from at least two sources of bias. The first results from our experimental manipulation in which the first 10 minutes in sessions of T1 and T2 were actually untreated. Given this nontreatment for the first 10 minutes, the aggregation based on the whole sample can result in an underestimate of the treatment effect. The second source of bias arises from the varying strength of treatment over time. We provide an illustrative example here to show the potential bias in treatment effect inferred from the mean comparison at the aggregate broadcaster level. For simplicity, we assume treatment strength is linear in time and each broadcaster only streams once. We describe the outcome variable y_{it} of broadcaster i at time t as follows:

$$y_{it}(Treat_i = 1) = \alpha + f(t) + \xi_i + \epsilon_{it}, \quad (1)$$

$$y_{it}(Treat_i = 0) = \alpha + \xi_i + \epsilon_{it}, \quad (2)$$

where $f(t)$ is the time-varying treatment effect and ξ_i denotes broadcaster heterogeneity with mean zero.

Figure 3. (Color online) Relationship Between the Number of Synthetic Viewers and Revenue Metrics



Notes. Each line represents the data fitted by a cubic spline and the surrounding areas represent the 95% confidence interval. The gap on the Y-axis in Figure 3(a) is caused by the discrete nature of the tip amount: the smallest tip is 10 cents (1 cent = 0.01 yuan), which causes the gap between log(10) = 2.30 and 0 on the Y-axis.

Based on Equation (1), the mean of y_{it} for broadcaster i from treatment groups is $\bar{y}_i = \left(\alpha + \sum_{t=1}^{T_i} f(t)/T_i + \xi_i + \sum_{t=1}^{T_i} \epsilon_{it}/T_i\right)$, where T_i represents the number of treated periods (session length). As we do not find a treatment effect on session length, we assume that T_i is independent and identically distributed, and therefore, we have

$$E_i(\bar{y}_i) = \alpha + E_{T_i}\left(\frac{\sum_{t=1}^{T_i} f(t)}{T_i}\right). \tag{3}$$

Table 5. Broadcaster-Level Mean Comparison Using the Whole Sample

	Mean			p-value		
	C	T1	T2	C vs. T1	T1 vs. T2	C vs. T2
Tip, cents	49.6	103.7	91.1	0.011*	0.341	0.052^{+}
TipRate, cents	4.51	6.57	4.87	0.049*	0.098^{+}	0.743
N^{real}	9.70	11.1	14.7	0.151	0.116	0.043*
Length, minutes	48.8	47.3	44.1	0.425	0.293	0.275
Number of sessions	17.7	13.8	14.8	0.108	0.358	0.175

Notes. p-value of one-sided t-test is reported. 1 cent = 0.01 yuan.
$^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.10$.

Consider a typical between-subjects design with multiple trials in which the treatment strength is constant (i.e., $f(t) = \beta$) across trials (i.e., minutes). In this scenario, we have $E(\bar{y}_i|Treat_i = 1) - E(\bar{y}_i|Treat_i = 0) = E_{T_i}\left(\sum_{t=1}^{T_i} f(t)/T_i\right) = \beta$, suggesting that the mean comparison estimate works. Nevertheless, the number of synthetic viewers added in our experiment accumulates over time, and therefore, the treatment strength is not identical across trials. If the treatment effect is perfectly linear in strength (i.e., $f(t) = \beta t$), we have $E(\bar{y}_i|Treat_i = 1) - E(\bar{y}_i|Treat_i = 0) = E_{T_i}\left(\sum_{t=1}^{T_i} f(t)/T_i\right) = \frac{E_{T_i}(T_i+1)}{2}\beta \propto \beta$, suggesting that the broadcaster-level mean comparison estimate is proportionally unbiased. However, if the treatment effect is nonlinear (e.g., $f(t) = \beta t + \gamma t^2$), we have $E(\bar{y}_i \mid Treat_i = 1) - E(\bar{y}_i \mid Treat_i = 0) = E_{T_i}\left(\frac{T_i+1}{2}\beta + \frac{(T_i+1)(2T_i+1)}{6}\gamma\right) = \frac{E_{T_i}(T_i+1)}{2}\beta + \frac{E_{T_i}\left[(T_i+1)(2T_i+1)\right]}{6}\gamma$, suggesting that the mean comparison estimate is no longer proportionally unbiased, and the extent of the bias is determined by the convexity of treatment effect and the distribution of session length.

To alleviate the first source of bias resulting from the nontreatment for the first 10 minutes, we rerun the mean comparison by excluding observations associated with the first 10 minutes from the two treatment groups. As the top panel of Table 6 shows, the means of variables of interest in the treatment groups all move toward the expected direction (e.g., *Tip* in T1 and T2 both move up compared with the values in Table 5). We cannot reject the hypothesis that *Tip* in C is smaller than those in T1 and T2 at the 5% level. In addition, *TipRate* significantly increases from C to T1 (p = 0.018) and then decreases from T1 to T2 (p = 0.034), suggesting an inverted U-shaped relationship. Furthermore, the differences in the number of real viewers between C and T1 and between T1 to T2 are statistically significant at the 10% level, suggesting an overall positive treatment effect on N^{real}. We also report the results of the mean comparisons in the bottom panel of Table 6 when excluding the first 10 minutes of data from both control and treatment groups. We find qualitatively similar patterns. To tackle the second source of bias resulting from varying treatment strength, we need to account for the variation in treatment strength over time and across sessions as we discuss next.

4.2. Slope Comparison

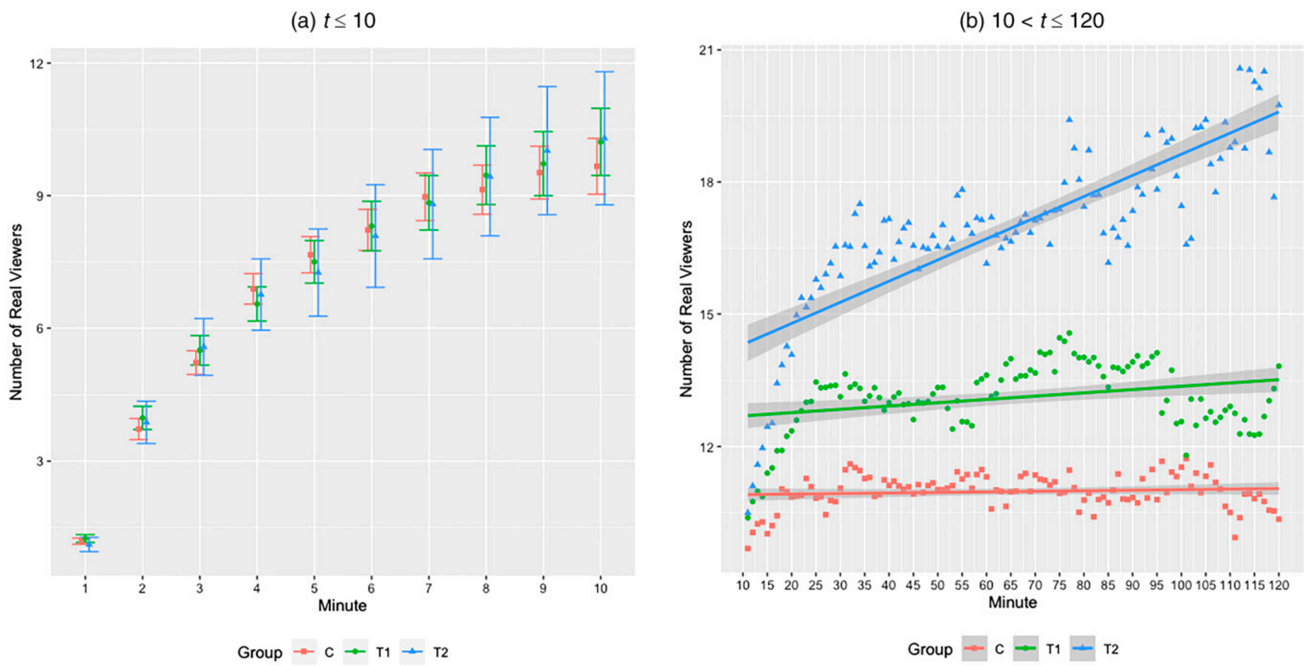
One intuitive approach of exploiting the variation in treatment strength over time is to compare the slope of variables of interest with time across treatment conditions. For example, if the addition of synthetic

Table 6. Broadcaster-Level Mean Comparison Excluding Observations with $t \leq 10$

	Mean			p-value		
	C	T1	T2	C vs. T1	T1 vs. T2	C vs. T2
			Excluding observations with $t \leq 10$ for T1 and T2			
Tip (cent)	49.6	128.7	95.4	0.005**	0.176	0.032*
TipRate (cent)	4.51	7.24	4.78	0.018*	0.034*	0.399
N^{real}	9.70	12.8	16.4	0.062^{+}	0.096^{+}	0.010*
			Excluding observations with $t \leq 10$ for T1, T2, and C			
Tip (cent)	57.6	128.7	95.4	0.012*	0.176	0.069^{+}
TipRate (cent)	4.93	7.24	4.78	0.042*	0.034*	0.558
N^{real}	9.99	12.8	16.4	0.094^{+}	0.096^{+}	0.016*

Notes. p-value of one-sided t-test is reported. 1 cent = 0.01 yuan.
$^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.10$.

Figure 4. (Color online) Dynamics of Number of Real Viewers Across Groups



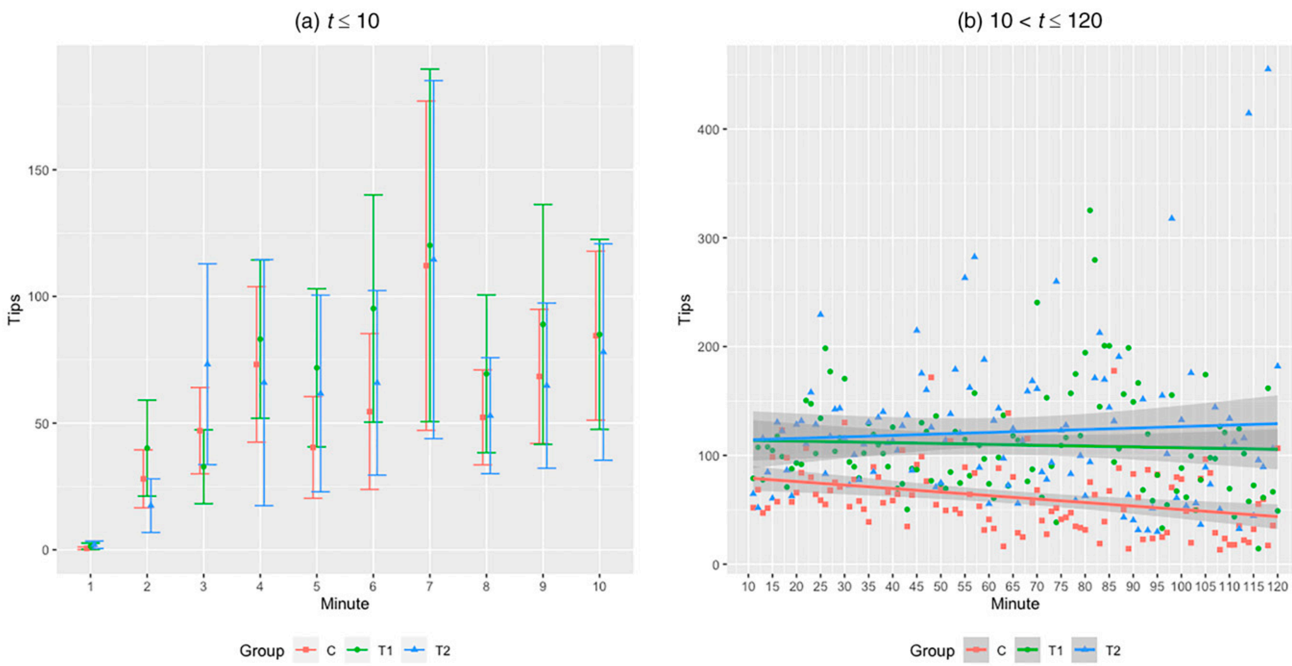
Note. In Figures 4–6, the minute-level mean estimates and error bars with 95% confidence interval are shown in the left figure, and the minute-level mean estimates and a linear fit line with 95% confidence interval are shown in the right figure.

viewers indeed increases the number of real viewers, we should expect the slope of N^{real} to be higher for T2 than T1 than C because the number of synthetic viewers on average increases faster over time in T2 than T1 than C. In addition, given that the addition of synthetic viewers proceeded after the 10th minute, we should not expect a significant difference in N^{real} across C, T1, and T2 during the first 10 minutes.

The dynamics of N^{real} over time confirm our expectations. As Figure 4(a) shows, the number of real viewers gradually increases during the first 10 minutes, but there is no substantial difference across the three groups. The p-values of the F-test that N^{real} is the same across the three groups at $t = 1, \ldots, 10$ have a mean of 0.559 and a minimum of 0.184. In addition, when $t \leq 10$, we are unable to reject the hypothesis that the slope in group C is the same as the slope in T1 and T2 ($p = 0.571$). These tests provide additional support for the success of our field experiment manipulations. By comparing the slope of N^{real} across groups when $10 < t \leq 120$, we find that the slope of T1 is steeper than C ($p < 0.001$) and the slope of T2 is steeper than T1 ($p < 0.001$), which is consistent with the positive treatment effect on N^{real}. As Figure 3 shows, the data with $N^{syn} > 1{,}200$ result from a few unusually long sessions, and $t = 120$ is approximately associated with the threshold of $N^{syn} = 1{,}200$ (i.e., the maximum of N^{syn} when $t \leq 120$ is 1,153). Thus, we use the data truncated at 120 minutes in subsequent analyses to prevent disproportionate influence from outliers.

We further plot the dynamics in *Tip* in Figure 5 and *TipRate* in Figure 6. Again, we are not able to reject the hypothesis that either *Tip* or *TipRate* is the same across C, T1, and T2 at $t = 1, \ldots, 10$ at the 5% level with the exception of *TipRate* at $t = 2$ ($p = 0.026$). By comparing the slopes during the nontreated period ($t \leq 10$), we cannot reject the hypothesis that the slope of *Tip* is the same across groups ($p = 0.670$), nor can we reject the same-slope hypothesis for *TipRate* ($p = 0.571$). During the treated period ($10 < t \leq 120$), we find that the slope of *Tip* in the control group is generally downward while the slope of *Tip* is either flat or slightly upward in two treatment groups, suggesting a positive treatment effect on *Tip*. Although the slopes for both T1 ($p = 0.028$) and T2 ($p = 0.012$) are significantly steeper than C, there is no significant difference in the slopes between T1 and T2 ($p = 0.552$). This data pattern again suggests that the treatment effect on tip amount is nonlinear and subject to diminishing returns. For the slope of *TipRate* shown in Figure 6(b), the downward slope for T1 is significantly less salient than C ($p < 0.001$) and significantly more salient than T2 ($p = 0.043$), and the slope for T2 is not significantly different from C ($p = 0.110$), which implies an inverted U-shaped relationship. Overall, the results from our slope comparisons provide additional support for the success of our manipulation in the experiment and confirm similar treatment effects on *Tip*, N^{real}, and *TipRate* found in previous mean comparisons with increased statistical power.

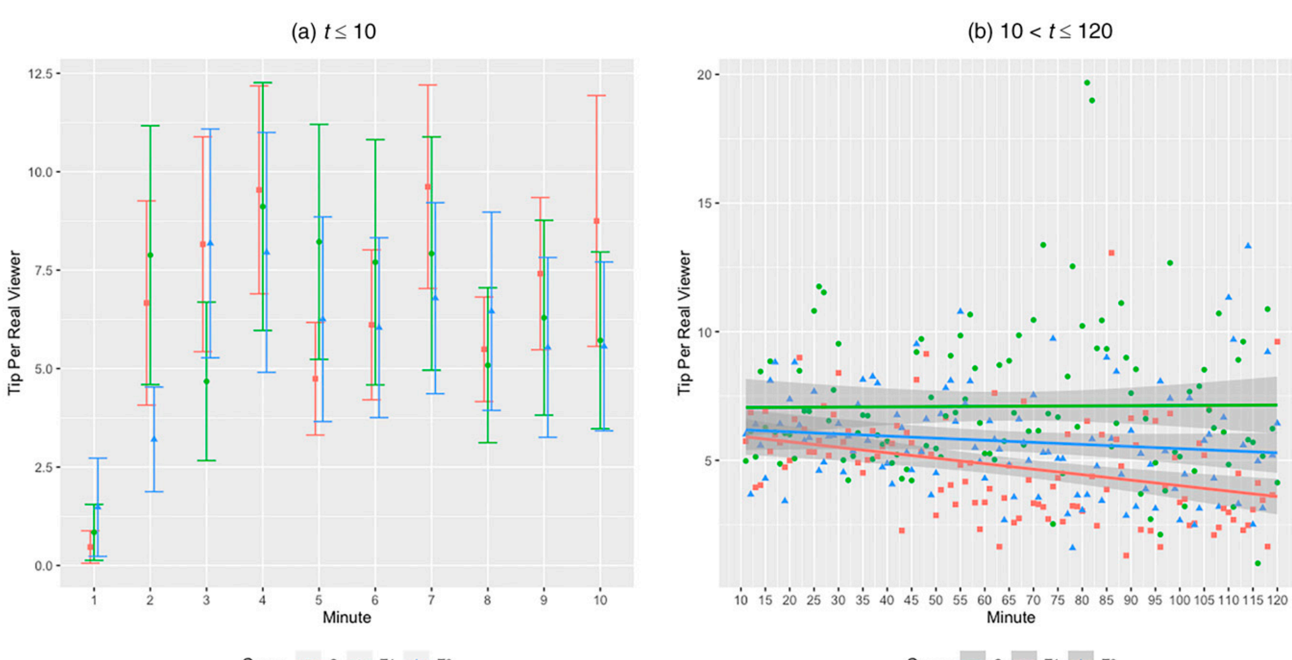
Figure 5. (Color online) Dynamics of Tip Amount Across Groups



4.3. Regression Analysis

A unique design of our field experiment is the randomization of synthetic viewers over time and across sessions in addition to the first level of randomization across broadcasters. This manipulation not only gives rise to the differential growth rates of synthetic viewers across groups (i.e., N^{syn} grows faster in T2 than T1 than C), but it also allows us to observe the exact treatment strength (i.e., N^{syn}) at each minute. This additional level of randomization enables us to exploit the richer temporal variation to identify the effect of audience size on tipping revenues.

Figure 6. (Color online) Dynamics of TipRate Across Groups



To estimate the causal impact of audience size, we employ a linear panel regression with session fixed effects. Compared with the mean comparison estimate, we are able to explicitly account for the variation in treatment strength in the regression, and therefore, the estimate is no longer subject to aggregation bias. An additional benefit of the regression analysis is that the coefficient estimate can provide more managerially relevant insights (e.g., the impact of increasing audience size by X on outcome Y), whereas the treatment effects inferred from between-group comparisons can only provide manipulation-specific insights (e.g., the impact of adding an average of two or four synthetic viewers per minute on outcome Y), which might not be of interest to other firms.

Let k denote a live streaming session and t denote a minute. The previous results regarding the effect of audience size on tipping revenue suggests a nonlinear relationship. To capture this potential nonlinear effect, we model the tip amount denoted by y_{kt} as a function of both the audience size and its squared term as follows:[6]

$$y_{kt} = \beta_1 N_{kt}^{disp} + \beta_2 \left(N_{kt}^{disp}\right)^2 + \beta_3 t + \beta_4 Tenure_{kt} + \eta_k + \epsilon_{kt}, \tag{4}$$

where N_{kt}^{disp} is the number of displayed viewers in session k at minute t; η_k represents the session-fixed effects, which account for any unobserved session-specific factors (e.g., quality of a session) that may affect viewers' tipping decisions; and ϵ_{kt} is the idiosyncratic error term.

We account for two different effects of timing in Equation (4). A broadcaster's level of engagement might change over time and, therefore, affect viewers' tipping behavior. Similarly, a viewer's motive to tip might also vary by the viewer's stay in a session. We account for the former by including the time since the beginning of a session denoted by t and account for the latter by including the average length of stay for real viewers in a session k at minute t denoted by $Tenure_{kt}$ in Equation (4). As our data are at the aggregate rather than the individual viewer level, we are unable to precisely measure how long each viewer remains in a session. Nevertheless, we create a proxy for the average length of stay for real viewers by making the following two assumptions:[7]

Assumption 1. *If $N_{kt}^{real} \geq N_{k,t-1}^{real}$, all existing real viewers do not exit and $(N_{kt}^{real} - N_{k,t-1}^{real})$ new real viewers arrive at time t. Thus, $Tenure_{kt} = Tenure_{k,t-1} \times N_{k,t-1}^{real} + N_{k,t-1}^{real}/N_{kt}^{real} = (Tenure_{k,t-1} + 1) \times N_{k,t-1}^{real}/N_{kt}^{real}$.*

Assumption 2. *If $N_{kt}^{real} < N_{k,t-1}^{real}$, $(N_{k,t-1}^{real} - N_{kt}^{real})$ existing real viewers exit and no new real viewers arrive at time t. Thus, $Tenure_{kt} = Tenure_{k,t-1} + 1$.*

Estimating Equation (4) directly might lead to biased estimates because N_{kt}^{disp} could be correlated with the error term ϵ_{kt} if there are unobserved factors (e.g., ambience) that affect both viewers' participation and tipping decisions. To address this potential endogeneity issue, we use N_{kt}^{syn} and its squared term as instrument variables (IVs) for N_{kt}^{disp} and its squared term. These are valid IVs because N_{kt}^{syn} is correlated with N_{kt}^{disp} by definition but not with the error terms as a computer algorithm exogenously generates N_{kt}^{syn}. Note that the audience size correlates with time trend as a result of the nature of our experiment design ($corr.\left(N_{kt}^{syn}, t\right) = .571$). However, N_{kt}^{syn} does not increase linearly over time because we added a *random* number of synthetic viewers rather than a fixed number to both T1 and T2. To employ the IV estimation, we focus on a subsample of our data set in which the IVs are well defined. In particular, we use observations from T1 and T2 when $10 < t \leq 120$ for model estimation because synthetic viewers were only added to treatment groups after the first 10 minutes. We also truncate at 120 minutes to mitigate the potential estimation bias caused by outliers.

We report estimation results of Equation (4) in Table 7, in which Model 1 includes the linear effect only and Model 2 includes potentially nonlinear effects. As we manipulated only the audience size in the experiment, we interpret the estimation results as the effect of audience size alone rather than a combined effect of the audience size and engagement activities of viewers. The significant and positive coefficients of N^{disp} in both models confirm the positive treatment effect of audience size on tipping revenues as we found previously. On average, increasing the audience size by one unit improves the tipping revenue per minute by approximately 0.01 yuan, which is 1% of the mean level. Such a positive effect of audience size on tipping revenue is subject to diminishing returns as indicated by the negative coefficient of the squared term (–0.00126, $p < 0.01$). The coefficient estimates from Model 2 suggest that the effect of audience size is mostly positive in our data range as it turns negative when N^{disp} exceeds $1.43/(2 \times 0.00126) = 567$, which is the 96.6 percentile in our data.

We next examine the drivers of the mostly positive effect of audience size on tipping revenue by breaking down the net impact on *Tip* to the impact on viewer participation (N^{real}) and average tip amount per viewer (*TipRate*) separately.

4.3.1. Does Adding Synthetic Viewers Draw More Real Viewers? We estimate the model of the number of real viewers as follows:

$$N_{kt}^{real} = \beta_1 N_{k,t-1}^{real} + \beta_2 N_{kt}^{syn} + \beta_3 \left(N_{kt}^{syn}\right)^2 + \beta_4 t + \beta_5 Tenure_{kt} + \eta_k + \epsilon_{kt}, \tag{5}$$

Table 7. Regression Results of Models of Tip, TipRate, and Number of Real Viewers

	DV = *Tip*		DV = *TipRate*		DV = N^{real}	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
N^{disp}	0.920**	1.43**	0.012*	0.022**		
	(0.215)	(0.296)	(0.005)	(0.008)		
$(N^{disp})^2$		−0.001**		−2.32e-5*		
		(3.34e-4)		(1.09e-5)		
Lag N^{real}					0.757**	0.754**
					(0.016)	(0.016)
N^{syn}					0.021**	0.028**
					(0.004)	(0.003)
$(N^{syn})^2$						−1.64e-5*
						(6.83e-6)
t	−2.57**	−3.11**	−0.057**	−0.067**	−0.040**	−0.045**
	(0.770)	(0.819)	(0.020)	(0.022)	(0.008)	(0.007)
Tenure	−4.97**	−5.40**	−0.103	−0.111	−0.238**	−0.246**
	(1.70)	(1.82)	(0.061)	(0.064)	(0.029)	(0.028)
Session fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Endogeneity correction	Yes	Yes	Yes	Yes		
Number of observations	53,893	53,893	53,893	53,893	52,833	52,833

Notes. Robust standard errors clustered by broadcasters in parentheses. Model 1 considers linear effect only; Model 2 includes both N and N^2.
$^{**}p < 0.01$; $^{*}p < 0.05$.

where we include the number of real viewers from the last period to capture the inertia in viewers' participation decisions. We also use N^{syn} rather than N^{disp} as the independent variable because $N^{disp} = N^{real} + N^{syn}$, suggesting that N^{disp}_{kt} perfectly correlates with ϵ_{kt} and, therefore, cannot serve as an independent variable.

As Table 7 shows, adding synthetic viewers draws more real viewers to live streaming sessions.[8] This is supported by the positive and significant coefficients of N^{syn} (0.021, $p < 0.01$), which suggests the existence of herding in a live streaming context. Adding 50 synthetic viewers results in approximately one additional real viewer. Although we find diminishing returns of the effect, the relatively small coefficient of $(N^{syn})^2$ suggests that the effect of audience size on viewer participation is mostly positive in our data range. That is, the effect is positive until N^{syn} exceeds $0.028/(2 \times 1.64\text{e-}5) = 854$, which is the 99.2 percentile in our data.

4.3.2. Does a Larger Audience Encourage or Discourage Average Tip Amount per Viewer?

As audience size increases, a viewer tends to obtain greater social image–related utility from tipping and, thus, might increase the tip amount. Nevertheless, a larger audience may also suggest greater competition for the broadcaster's reciprocal acts, which results in a lower marginal utility of tipping and further reduces the willingness to tip. To empirically test the direction of the effect, we estimate the model of *TipRate* using Equation (4) with $y_{kt} = TipRate_{kt}$. We report the results from the IV estimation in Table 7 under the column DV = *TipRate*. The coefficient estimates reveal an overall positive relationship between the audience size and *TipRate*, suggesting the dominance of social image–related utility over seeking reciprocity in driving viewers' tipping when N^{disp} is not too large. When N^{disp} exceeds a certain threshold ($0.022/[2 \times 2.32\text{e-}5] = 474$, which is the 94.2 percentile), the relationship reverses, perhaps because of the dominance of the negative competition effect.

5. Model Extensions and Robustness Checks

We first extend our main analysis by further exploring the heterogeneity in the treatment effect of increasing audience size. We then conduct a series of robustness checks to show that our findings are not sensitive to alternative model specifications and assumptions.

5.1. The Moderating Effect of Broadcaster Characteristics

5.1.1. Does the Effect of Audience Size Differ by Broadcaster Gender?

Previous research suggests that herding occurs if a viewer refers to existing viewers' choices as a descriptive social norm and, thus, passively mimics their decisions to follow well-attended sessions (Croson and Shang 2008, Simonsohn and Ariely 2008). Our pretreatment data show that live streaming sessions by female broadcasters are better attended than sessions by male broadcasters as indicated by the significant difference (8.00, $p < 0.01$) in the number of chatters per session between female and male broadcasters. The theory of social norm suggests that a viewer's choice of a mainstream product is more justifiable than the choice of a niche product,

Table 8. Regression Results on Subsamples (Female vs. Male Broadcasters)

	Female broadcasters only			Male broadcasters only		
	DV = *Tip*	DV = *TipRate*	DV = N^{real}	DV = *Tip*	DV = *TipRate*	DV = N^{real}
N^{disp}	1.55**	0.025**		1.01*	0.011	
	(0.328)	(0.009)		(0.450)	(0.019)	
$(N^{disp})^2$	−0.001**	−2.19e-5*		−0.001	−4.07e-5	
	(3.50e-4)	(9.60e-6)		(0.001)	(3.36e-5)	
Lag N^{real}			0.741**			0.706**
			(0.017)			(0.017)
N^{syn}			0.032**			0.016**
			(0.003)			(0.004)
$(N^{syn})^2$			−1.55e-5*			−1.96e-5*
			(6.43e-6)			(7.72e-6)
t	−3.20**	−0.074**	−0.048**	−2.35*	−0.028	−0.026**
	(0.962)	(0.025)	(0.008)	(1.10)	(0.042)	(0.005)
Tenure	−4.81**	−0.064	−0.284**	−5.71	−0.196	−0.154**
	(2.29)	(0.076)	(0.035)	(2.94)	(0.116)	(0.044)
Session fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Endogeneity correction	Yes	Yes		Yes	Yes	
Number of observations	38,613	38,613	37,944	15,280	15,280	14,889

Note. Robust standard errors clustered by broadcasters in parentheses.
$^{**}p < 0.01$; $^{*}p < 0.05$.

and thus, other viewers are more likely to follow this viewer to adopt the mainstream rather than the niche product. Following this reasoning, as female broadcasters are the mainstream on our platform, we expect the effect of popularity information on viewer participation to be stronger in sessions by female broadcasters than male broadcasters.

Nevertheless, previous research also indicates that consumers often engage in observational learning to draw quality inference from the observation of others' decisions (e.g., Banerjee 1992, Zhang and Liu 2012). For example, Tucker and Zhang (2011) show that niche products with narrow appeal benefit more from popularity information than broad-appeal products in e-commerce because consumers infer higher quality from a narrow-appeal product than an equally popular broad-appeal product. Given the majority of male viewers and female broadcasters, we can classify sessions by male broadcasters as niche products. Thus, if observational learning drives herding in our data, male broadcasters benefit more than female broadcasters from the increase in popularity in drawing viewers.

Broadcaster gender may also moderate an individual's motivation for tipping. Theories from evolutionary psychology suggest that men are likely to display financial resources as a tactic to signal status and gain attention from female mates, which results in a higher level of conspicuous consumption by men than women (Buss 1988, Griskevicius et al. 2007, Kenrick and Griskevicius 2013). In our context of live streaming, the theory of intrasexual competition predicts that such status-signaling motivation among male customers is more likely to appear in the presence of a female rather than a male broadcaster (Sundie et al. 2011). Given the male-dominant audience on our platform, we expect the utility related to social status and prestige to be greater in sessions by female than male broadcasters, which predicts a stronger effect of audience size on average tip amount (a form of conspicuous consumption) in female broadcasters' sessions.

To test the moderating effect of broadcaster gender, we divide the sample into female versus male broadcasters and reestimate Equations (4) and (5). Table 8 reports the results. Although the main effects of audience size on *Tip* are both significant and positive for female and male broadcasters, the coefficient of N^{disp} is smaller for male broadcasters. Next, we compare the effect of audience size on N^{real} and *TipRate* between female and male broadcasters, respectively. We find that the coefficient of N^{syn} is larger in the model of N^{real} for female broadcasters (0.032, $p < 0.01$) than for male broadcasters (0.016, $p < 0.01$). This result provides suggestive evidence that social norm rather than observational learning drives the herding in viewers' participation in live streaming. Regarding the regression results of *TipRate*, we find only a significant effect of N^{disp} for female broadcasters (0.025, $p < 0.01$) although the effect is not statistically significant for male broadcasters (0.011, $p > 0.05$). This result implies that the motivation of signaling social status only exists in female broadcasters' sessions, which is consistent with the theory of intrasexual competition. We show in Online Appendix C that the differential effects of audience size on *Tip*, *TipRate*, and N^{real} between female and male broadcasters still hold when we test the moderating effect using interaction terms between audience size and gender indicator variable.

5.1.2. Does the Effect of Audience Size Vary by Broadcaster Fame? Viewers on the live streaming platform do not know all broadcasters equally. From a viewer's perspective, tipping a star broadcaster rather than a nonstar broadcaster can improve the viewer's social image because of the perceived affiliation with a more prestigious group in front of other viewers, holding everything else equal. Thus, we might expect a greater effect of audience size on tipping revenue in sessions by more famous broadcasters.

We operationalize the fame of each broadcaster by the dollar value of all tips received during the one-month pretreatment period because the platform publicizes the rank of broadcasters on the basis of their monthly earnings with the top broadcaster earning the most. The variable $Fame_i$, defined as the monthly tip amount, has a mean of 403.1 yuan and a maximum of 6,600 yuan in our sample, suggesting that none of the broadcasters are superstars in the conventional sense. Nevertheless, we still have a large variation in the monthly earnings of broadcasters (SD = 809.7), which enables us to identify the moderating effect of fame.

We add interaction terms $N \times Fame_i$ and $N^2 \times Fame_i$ to Equations (4) and (5) and report the regression results in Table 9. The coefficient of $N^{disp} \times Fame$ is significant and positive in the regression of Tip, suggesting a stronger effect of audience size in sessions by more famous broadcasters. The most earned broadcaster benefits approximately 180% (4.03e-4 × [6,600 − 403]/1.40 = 1.78) more than an average broadcaster from the increase in audience size. The coefficient of $N^{disp} \times Fame$ in the regression of $TipRate$ is positive but statistically insignificant, which does not support the enhanced social image from tipping to top-ranked broadcasters. The positive moderation of $Fame$ in the effect of audience size on Tip seems to be driven by the stronger effect on viewer participation as suggested by the positive and significant coefficient of $N^{fake} \times Fame$ in the regression of N^{real}. As sessions by more famous broadcasters' exhibit more viewers, this result is consistent with our previous finding that social norm rather than observational learning drives herding on our platform.

Table 9. The Moderating Effect of Broadcaster Fame

	DV = Tip	DV = $TipRate$	DV = N^{real}
N^{disp}	1.40**	0.021**	
	(0.293)	(0.008)	
$(N^{disp})^2$	−0.001**	−2.23e-5*	
	(2.73e-4)	(1.04e-5)	
$N^{disp} \times Fame$	4.03e-4**	3.87e-6	
	(1.17e-4)	(2.66e-6)	
$(N^{disp})^2 \times Fame$	−2.73e-7	−2.68e-9	
	(1.85e-7)	(6.80e-9)	
$Lag\ N^{real}$			0.709**
			(0.024)
N^{syn}			0.032**
			(0.003)
$(N^{syn})^2$			−1.70e-5**
			(5.58e-6)
$N^{syn} \times Fame$			7.03e-6**
			(1.75e-6)
$(N^{syn})^2 \times Fame$			1.89e-9
			(7.19e-9)
t	−3.01**	−0.066**	−0.054**
	(0.863)	(0.022)	(0.008)
$Tenure$	−5.19**	−0.109	−0.275**
	(1.76)	(0.065)	(0.033)
Session fixed effects	Yes	Yes	Yes
Endogeneity correction	Yes	Yes	
Number of observations	53,893	53,893	52,833

Notes. Robust standard errors clustered by broadcasters in parentheses. N^{disp}, N^{syn}, and $Fame$ are mean-centered.
$^{**}p < 0.01$; $^{*}p < 0.05$.

5.2. The Moderating Effect of Broadcaster Performances

We further investigate the potential moderating effect of broadcaster performances. As time-variant performance metrics are potentially endogenous to the treatment, the results derived from this model extension are better interpreted as exploratory rather than conclusive. Our main analysis reveals a positive causal impact of audience size on viewers' tipping when the audience size is not too large, which is consistent with the existence of social image–related utility in tipping. In a live steaming session, a broadcaster's recognition of a viewer's action would likely garner attention from other viewers, thus enhancing the social image–related utility of the focal viewer. For example, a broadcaster may welcome a viewer when this viewer joins a session or express gratitude to the viewer for the tip received. In line with this logic, a viewer is more likely to tip in front of other viewers in anticipation of a broadcaster's recognition.[9] We, therefore, expect a broadcaster's recognition tendency to positively moderate the effect of audience size on viewers' tipping.

To test this hypothesis, we operationalize a broadcaster's recognition tendency by the frequency of saying words, as suggested by the platform, related to greeting and gratitude toward a specific viewer during a session: "welcome," "hello," "thank you," and "many thanks" (in Chinese). We create $Recognition_{kt}$, the count of these four words by the broadcaster in session k during minute t, by employing the state-of-the-art speech recognition algorithms detailed in Online Appendix B.

We also create three other metrics (i.e., $FaceTime_{kt}$, $HandTime_{kt}$, and $Emotion_{kt}$) from the video content as proxies of the intensity and quality of the broadcaster's performance (for definitions and summary statistics, see Table 3). Specifically, we measure the intensity of performance according to the percentage of frames in which the broadcaster's face and hands appear in front of the screen within a minute. Intuitively, the longer a broadcaster disappears from the screen, the lower the performance intensity perceived by viewers. We assume that the quality of a session is associated with a broadcaster's emotion, which we infer from facial expressions using Microsoft Azure Emotion. This program detects and analyzes facial expressions using deep convolutional neural networks. The output of these algorithms is a vector of scores associated with eight types of positive and negative emotions (see Online Appendix B). We aggregate these scores to create an emotion score, $Emotion_{it}$, to measure the extent of positivity in a broadcaster's facial expression within a minute.

To evaluate the moderating effects, we add interaction terms $N \times Recognition_{kt}$ and $N^2 \times Recognition_{kt}$ to Equations (4) and (5). We also include $Recognition_{kt}$ as an independent variable because of the variation of $Recognition_{kt}$ over time and across sessions, which session fixed effects do not capture. We present estimation results in Table 10. The results show that the positive impact of audience size on *Tip* is stronger in a session in which the broadcaster is more likely to express recognition. Broadcasters whose frequency of using recognition-related words is one standard deviation above population mean enjoy 10.0% (6 × 0.023/1.38 = 0.100) more revenue from the increase in audience size. The stronger effect on *Tip* results from the moderating effect of *Recognition* on *TipRate* as indicated by the positive and significant coefficient of $N^{disp} \times Recognition_{kt}$ in the model of *TipRate*. These findings again provide evidence for the existence of social image–related utility in tipping.

Table 10. The Moderating Effect of Broadcaster's Recognition Tendency

	DV = *Tip*	DV = *TipRate*	DV = N^{real}
N^{disp}	1.38**	0.021**	
	(0.295)	(0.008)	
$(N^{disp})^2$	–0.001**	–2.35e-5*	
	(2.90e-4)	(1.16e-5)	
$N^{disp} \times Recognition$	0.023*	7.99e-4*	
	(0.010)	(3.22e-4)	
$(N^{disp})^2 \times Recognition$	−1.22e-5	−4.61e-7	
	(3.38e-5)	(8.22e-7)	
Lag N^{real}			0.726**
			(0.021)
N^{syn}			0.031**
			(0.004)
$(N^{syn})^2$			−2.11e-5**
			(6.38e-6)
$N^{syn} \times Recognition$			−1.44e-4
			(1.51e-4)
$(N^{syn})^2 \times Recognition$			1.23e-6**
			(3.22e-7)
Recognition	0.069	–0.021	3.52e-4
	(1.04)	(0.026)	(0.005)
t	−2.98**	–0.065**	–0.051**
	(0.814)	(0.022)	(0.009)
Tenure	–4.67**	–0.084	–0.275**
	(1.64)	(0.059)	(0.034)
Session fixed effects	Yes	Yes	Yes
Endogeneity correction	Yes	Yes	
Number of observations	53,893	53,893	52,833

Notes. Robust standard errors clustered by broadcasters in parentheses. N^{disp}, N^{syn}, and *Recognition* are mean-centered.
$^{**}p < 0.01$; $^{*}p < 0.05$.

Further analyses provide additional evidence in support of our proposed mechanism through the social image–related utility. We examine whether broadcasters' behaviors (i.e., $FaceTime_{kt}$, $HandTime_{kt}$, $Emotion_{kt}$) that are unlikely to be related to the signaling value of a viewer's tipping also moderate this positive effect. Note that these types of broadcaster behaviors differ from $Recognition_{kt}$ as a broadcaster's body language and facial expressions are often made to the general audience rather than to a particular viewer. Therefore, we do not expect $FaceTime_{kt}$, $HandTime_{kt}$, or $Emotion_{kt}$ to moderate the impact of N_{kt}^{disp} on social image–related utility of tipping; thus, we propose a falsification test. To conduct the proposed falsification test, we reestimate the model of *Tip* in a similar way as we did for the test for *Recognition*. The results in Table 11 confirm that neither of these three performance metrics moderates the effect of audience size on tipping revenue. Aside from broadcaster performances, we explore a few other potential moderators in Online Appendix C and find that past tips and the number of real viewers moderate the effect of audience size on tipping, but viewer engagement does not moderate this effect.

5.3. Robustness Checks

We conduct a series of analyses to ensure the robustness of our main findings in Online Appendix D. We first show that the positive effect of audience size on tipping revenue still holds after controlling for potential confounding factors, including the proportion of synthetic viewers, the proportion of new viewers, and the potential spillover effects of other concurrent streams.

We then consider an alternative model specification without session fixed effects but with the time-of-day and social norm effects. We still find a mostly positive effect of audience size on each of the three outcome variables (*Tip*, *TipRate*, N^{real}). In addition, total tips tend to be higher in sessions by female and famous broadcasters. Viewers tend to tip more in talk shows than other genres. Regarding the timing of broadcasting, sessions that start in the early morning (8 a.m.) and late afternoon (4 p.m.) receive significantly fewer tips than sessions that start at midnight.

We exclude all short sessions that last no more than 10 minutes in our regressions because of the lack of variation in the number of synthetic viewers during the first 10 minutes. In total, 482 out of 2,222 sessions are short sessions and contribute to 0.47% of total

Table 11. The Moderating Effect of Additional Performance Metrics of Broadcasters

	DV = *Tip*		
	M = *FaceTime*	M = *HandTime*	M = *Emotion*
N^{disp}	1.40**	1.43**	1.42**
	(0.282)	(0.287)	(0.301)
$(N^{disp})^2$	−0.001**	−0.001**	−0.001**
	(3.20e-4)	(2.90e-4)	(3.51e-4)
$N^{disp} \times M$	0.377	0.084	0.517
	(0.268)	(0.326)	(0.461)
$(N^{disp})^2 \times M$	2.20e-4	−2.10e-4	−0.001
	(4.22e-4)	(7.75e-4)	(0.001)
M	−19.4	7.75	8.44
	(19.9)	(19.0)	(35.4)
t	−3.11**	−3.12**	−3.10**
	(0.862)	(0.844)	(0.823)
Tenure	−5.44**	−5.43**	−5.43**
	(1.91)	(1.82)	(1.81)
Session fixed effects	Yes	Yes	Yes
Endogeneity correction	Yes	Yes	Yes
Number of observations	53,893	53,893	53,893

Notes. Robust standard errors clustered by broadcasters in parentheses. N^{disp}, *FaceTime, HandTime*, and *Emotion* are mean-centered.
$^{**}p < 0.01$; $^{*}p < 0.05$.

revenues. The exclusion of short sessions might give rise to a sample selection problem if the social benefits of tipping systematically differ between short and long sessions. We provide evidence that our findings do not seem to be subject to the sample selection bias.

We not only observe data across the three treatment groups but also have data before and after the experiment, which allows us to check the robustness of findings using a difference-in-differences (DID) analysis. We present corroborating evidence of the positive and nonlinear treatment effect on tipping revenue from the estimation of a DID model.

To further test the proposed behavioral mechanism through social image, we investigate the effects of audience size on chats and likes as a falsification test. We find a negative main effect of audience size on *ChatRate* and a null effect on *LikeRate,* which are consistent with the absence of social image in these nonmonetary engagement activities. One possible explanation for the main negative effect on *ChatRate* is the substitution between tipping and chatting. When audience size grows, the increasing social image–related utility in tipping may drive viewers to switch from chatting to tipping, which explains the downward trend in *ChatRate* against audience size. Notably, as nonmonetary engagement activities, such as chats and likes, are often used as key health indicators by online communities and platforms, the decrease in chats against audience size could be a potential downside of the PWYW model.

We also empirically rule out an alternative mechanism that the positive effect of audience size on tipping is through changes in service quality manifested in broadcaster behavior. Specifically, none of the four lag broadcaster performance metrics has a statistically significant relationship with either *Tip* or *TipRate*, suggesting that broadcaster behavior does not mediate the positive main effect of audience size on tipping. For broadcaster performance, we find that a larger audience size generally increases the duration of a broadcaster's face appearing in front of the camera (i.e., *FaceTime*). This finding suggests that broadcasters tend to increase their broadcasting effort when having a larger audience. Although we did not find a significant relationship between *FaceTime* and viewers' immediate tips, an increased frequency of showing faces to viewers might improve the viewer-broadcaster relationship in the long run and, therefore, lead to a greater scalability of PWYW. Given the relatively short span of our data, we are unable to test this hypothesis in this study. We, therefore, leave these interesting questions pertaining to the long-run benefits of a large audience to future research.

Finally, we conduct a descriptive analysis by examining the correlations between audience size and key outcome variables using data from the control group alone, in which all viewers are real. Consistent with our main findings from the experiment, we observe a mostly positive relationship between audience size (N^{disp}) and total tips (*Tip*) and a mostly positive relationship between audience size and average tip per real viewer (*TipRate*). These observations suggest that not only having more synthetic viewers, but also having more regular real viewers, may improve tipping revenue.

6. Discussion and Conclusion

Fueled by the advances in mobile technology and internet penetration, the live streaming industry has grown to be a billion-dollar market and has become a viable source of income for thousands of people across the globe. As major players in the social media industry embrace live streaming, the question regarding the scalability of the PWYW-based revenue model (i.e., whether larger audiences generate greater revenues) remains unanswered.

We find a mostly positive casual effect of audience size on tipping revenues using data from a field experiment run on a live streaming platform that uses PWYW. However, the effect of audience size is concave, and it can even turn negative when audience size is relatively large. By decomposing the net effect on revenue to the effect on viewer size and on tip per viewer, we find that the positive effect on total revenue results from a positive effect of popularity information on viewer size and a mostly positive effect on individual payment. The former finding is consistent with theories related to herding; the latter finding suggests the dominance of social image motive over seeking reciprocity in driving a viewer's tipping in live streaming. We also find that famous female broadcasters who use recognition-related words frequently benefit the most from an increase in audiences. In terms of mechanisms, we provide suggestive evidence for the importance of social image–related utility in PWYW through moderation analyses and falsification tests. We also scrutinize the possibility that audience size can improve the performance quality of broadcasters and, thus, have an indirect effect on tipping. The lack of a statistically significant relationship between broadcasters' recent performances and tipping suggests the absence of such an indirect effect.

Our research adds to the PWYW literature by providing the first assessment of the scalability of revenues generated by PWYW, which sheds light on the viability of PWYW as a key monetization tool for firms, especially for social media firms that rely on UGC. In addition, we complement previous PWYW studies by showing that an important motive for consumers to pay under PWYW is to improve their social images, at least in the live streaming context. We also contribute to the burgeoning live streaming literature by presenting one of the first few studies to use field rather than survey data.

Our findings provide several important implications to marketing practitioners. Because of the concave relationship between audience size and tipping revenue, platforms might consider splitting viewers into multiple chat rooms to alleviate their declining marginal benefit of tipping, especially for extremely populous sessions. The validity of this recommendation hinges on the assumption that either the tipping rate does not depend on the broadcaster' performance quality or the broadcasters' performance quality does not depend on the audience size. When this assumption does not hold, splitting viewers into smaller groups might lower the broadcaster's performance quality, which, in turn, lowers the tipping amount. Although we find support for this assumption in our empirical data, we caution firms that they should scrutinize this assumption before following our recommendation of splitting viewers.

A caveat of using PWYW in live streaming is that it does not scale linearly with the viewership as other monetization methods do, such as advertising and subscription. Thus, a live streaming firm might consider diversifying its monetization methods to accommodate advertising and/or subscriptions when it reaches a certain stage. Although we used synthetic viewers as a part of the research design, we do not recommend platforms to add synthetic viewers to live streaming sessions to increase revenues because of the unethical nature of this practice. Instead, given the importance of social image–related utility in PWYW, live streaming platforms could offer additional status-seeking devices, such as badges, avatars, or ranking systems, to further improve revenues. We also caution firms that intentionally manipulating audience size may backfire because the revenue will actually decrease rather than increase when the number of displayed viewers is relatively large.

We focus on the context of live streaming in this research. However, the question pertaining to scalability extends to other important formats of virtual communication, such as massive open online courses (MOOC). As consumers are increasingly migrating from the traditional off-line communication to the "new normal" of online/virtual communication, it is imperative for firms to understand and quantify the marginal benefit of expanding the audience given that the marginal cost of holding a larger audience is negligible in the virtual space. We take the first step in this research to examine the scalability of PWYW as a business model for live streaming. Future research could extend our research design to explore the scalability of other forms of virtual communication. For example, it will be interesting to study how the overall learning effectiveness of a MOOC might vary with the number of enrolled students.

Our study has limitations, which suggest avenues for future research. One limitation of this study is the relatively small sample size; thus, larger sample size exploration should help establish robustness. In addition, we do not observe individual-level data and, therefore, are unable to assess the potential heterogeneity in treatment effects across viewers. For example, viewers who tipped a lot in the past might gain more benefit from broadcasters' attention than those marginal viewers who rarely tipped. Should individual-level data be available, researchers could extend our analysis to investigate heterogeneous treatment effects across viewers. We started the treatment after the first 10 minutes, which might give rise to the sample selection problem. Although we find no strong evidence of sample selection in our empirical context, future research could improve our experiment design to avoid the sample selection a priori. Finally, we manipulated only the audience size but not any engagement activity (e.g., tips, chats, likes) in this experiment. An investigation of the separate effects of different types of engagement activities on revenue could be a fruitful direction for future research.

Acknowledgments
The authors thank an anonymous live streaming platform in China for implementing the study and providing the data. The authors also grateful to the editor, associate editor, and two anonymous reviewers for their constructive guidance. The authors also thank the participants at the 2017 Marketing Dynamics Conference; the 2017 Marketing Science Conference; the 2017 NYU-Temple Conference on Digital, Mobile Marketing, and Social Media Analytics; the 2018 China India Insights Conference; the 2018 NUS-Tsinghua Workshop on Digital Economy; the 2018 UTD-FORMS Conference; Hong Kong University, Peking University, Fudan University, Xiamen University, University of Texas at Austin, and University of North Carolina at Chapel Hill for their valuable feedback.

Endnotes
[1] Lin et al. (2021) also use the field data to study drivers of tipping. Different from our focus on audience size, they focus on the interactions between broadcasters' and viewers' emotions.

[2] According to Lin and Lu (2017), the percentage of female broadcasters and male viewers from our platform are close to the industry average (73% female broadcasters and 75% male viewers).

[3] The exchange rate between Chinese yuan and U.S. dollars is around 6.6 yuan per dollar as of August 1, 2016.

[4] The platform did not provide us the number of real viewers during the pretreatment period because it was unable to retrieve this variable from the database for some technical reasons.

[5] Mainly because of the company's data policy, the sample size collected in this research is more moderate than those in recent field experiments in marketing (e.g., Sudhir et al. 2016, Dubé et al. 2017). Nevertheless, we observe granular data at the minute level within sessions.

[6] We also test a linear-log model in which N_{kt}^{disp} is log-transformed. The smaller mean absolute error of the simple linear model with squared term (100.30) than that of the linear-log model (101.53) supports our model specification.

[7] The variable *Tenure* has a mean of 9.18 and a standard deviation of 5.34 under these two assumptions.

[8] Because of the inclusion of $N_{k,t-1}^{real}$, we also estimated Equation (5) using the Arellano–Bond estimator (Arellano and Bond 1991) and found statistically equivalent results.

[9] A broadcaster with a high tendency to recognize viewers does not necessarily interact with a viewer more frequently because the recognition tendency can be driven by a broadcaster's politeness rather than reciprocity.

References
Andreoni J (1988) Privately provided public goods in a large economy: The limits of altruism. *J. Public Econom.* 35(1):57–73.
Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econom. Stud.* 58(2):277–297.
Ariely D, Bracha A, Meier S (2009) Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *Amer. Econom. Rev.* 99(1):544–555.
Azar OH (2007) The social norm of tipping: A review. *J. Appl. Soc. Psych.* 37(2):380–402.
Azar OH (2011) Business strategy and the social norm of tipping. *J. Econom. Psych.* 32(3):515–525.
Banerjee AV (1992) A simple model of herd behavior. *Quart. J. Econom.* 107(3):797–817.
Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Political Econom.* 100(5):992–1026.
Buss DM (1988) The evolution of human intrasexual competition: Tactics of mate attraction. *J. Personality Soc. Psych.* 54(4):616–618.
Cai H, Chen Y, Fang H (2009) Observational learning: Evidence from a randomized natural field experiment. *Amer. Econom. Rev.* 99(3):864–882.
Chen Y, Koenigsberg O, Zhang ZJ (2017) Pay-as-you-wish pricing. *Marketing Sci.* 36(5):780–791.
China Daily (2016) China's internet celebrity economy bigger than cinema. Accessed May 25, 2021, http://www.chinadaily.com.cn/china/2016-09/17/content_26812402.htm.
Croson R, Shang JY (2008) The impact of downward social information on contribution decisions. *Experiment. Econom.* 11(3):221–233.
DellaVigna S, List JA, Malmendier U (2012) Testing for altruism and social pressure in charitable giving. *Quart. J. Econom.* 127(1):1–56.
Deloitte (2018) Technology, media and telecommunications predictions 2018. Accessed May 25, 2021, https://www2.deloitte.com/cn/en/pages/technology-media-and-telecommunications/articles/tmt-predictions-2018.html.
Dubé J-P, Luo X, Fang Z (2017) Self-signaling and prosocial behavior: A cause marketing mobile field experiment. *Marketing Sci.* 36(2):161–186.
D'Urbino L (2017) China's new craze for live-streaming. A new way of bringing colour to dreary lives. *The Economist Online* (February 9), https://www.economist.com/special-report/2017/02/09/chinas-new-craze-for-live-streaming.
Garcia SM, Tor A (2009) The N-effect: More competitors, less competition. *Psych. Sci.* 20(7):871–877.
Gneezy A, Gneezy U, Nelson LD, Brown A (2010) Shared social responsibility: A field experiment in pay-what-you-want pricing and charitable giving. *Sci.* 329(5989):325–327.
Gneezy A, Gneezy U, Riener G, Nelson LD (2012) Pay-what-you-want, identity, and self-signaling in markets. *Proc. Natl. Acad. Sci. USA* 109(19):7236–7240.
Griskevicius V, Tybur JM, Sundie JM, Cialdini RB, Miller GF, Kenrick DT (2007) Blatant benevolence and conspicuous consumption: When romantic motives elicit strategic costly signals. *J. Personality Soc. Psych.* 93(1):85–102.
Harbaugh WT (1998) The prestige motive for making charitable transfers. *Amer. Econom. Rev.* 88(2):277–282.
Hays S (2018) Esports: The "tipping market"—$129,000,000 in tips to streamers paid in 2017. Accessed May 25, 2021, https://medium.com/hackernoon/esports-the-tipping-market-129-000-000-in-tips-to-streamers-paid-in-2017-2cd7248ee623.

He Q, Liu J, Wang C, Li B (2016) Coping with heterogeneous video contributors and viewers in crowdsourced live streaming: A cloud-based approach. *IEEE Trans. Multimedia* 18(5):916–928.

iiMedia Research (2019) 2018-2019 China online live broadcasting industry research report. Accessed May 25, 2021, https://www.iimedia.cn/c400/63478.html.

iiMedia Research (2020) 2019-2020 China online live streaming market research report. Accessed May 25, 2021, https://www.iimedia.cn/c400/69017.html.

Jung MH, Perfecto H, Nelson LD (2016) Anchoring in payment: Evaluating a judgmental heuristic in field experimental settings. *J. Marketing Res.* 53(3):354–368.

Jung MH, Nelson LD, Gneezy A, Gneezy U (2017) Signaling virtue: Charitable behavior under consumer elective pricing. *Marketing Sci.* 36(2):187–194.

Kenrick DT, Griskevicius V (2013) *The Rational Animal: How Evolution Made Us Smarter Than We Think* (Basic Books, New York).

Kim JY, Natter M, Spann M (2009) Pay what you want: A new participative pricing mechanism. *J. Marketing* 73(1):44–58.

Knowledge@Wharton (2015) Live—and lucrative? Why video streaming supremacy matters. Accessed May 25, 2021, https://knowledge.wharton.upenn.edu/article/live-and-lucrative-why-video-streaming-supremacy-matters/.

Lampel J, Bhalla A (2007) The role of status seeking in online communities: Giving the gift of experience. *J. Comput. Mediated Comm.* 12(2):434–455.

Lee YC, Yen CH, Wang D, Fu WT (2019) Understanding how digital gifting influences social interaction on live streams. *Proc. 21st Internat. Conf. Human Comput. Interaction Mobile Devices Services* (ACM, Taipei, Taiwan), 1–10.

Lin J, Lu Z (2017) The rise and proliferation of live-streaming in China: Insights and lessons. *Internat. Conf. Human Comput. Interaction* (Springer, Cham, Switzerland), 632–637.

Lin Y, Yao D, Chen X (2021) Happiness begets money: Emotion and engagement in live streaming. *J. Marketing Res.* 58(3):417–438.

Mediakix (2018) The top 16 Facebook live statistics you should know. Accessed May 25, 2021, https://mediakix.com/blog/facebook-live-statistics-video-streaming-to-know/#gs.vR8kVHA.

Natter M, Kaufmann K (2015) Voluntary market payments: Underlying motives, success drivers and success potentials. *J. Behav. Experiment. Econom.* 57:149–157.

Olson M (1965) *Logic of Collective Action: Public Goods and the Theory of Groups* (Harvard University Press, Cambridge, MA).

Pires K, Simon G (2015) YouTube live and Twitch: A tour of user-generated live streaming systems. *Proc. 6th ACM Multimedia Systems Conf.* (ACM, Portland, OR), 225–230.

Qin A (2016) China's viral idol: Papi Jiang, a girl next door with attitude. *The New York Time Online* (August 24), https://www.nytimes.com/2016/08/25/arts/international/chinas-viral-idol-papi-jiang-a-girl-next-door-with-attitude.html.

Roettgers J (2018) Facebook adds ability to tip live streamers to mobile apps. *Variety Online* (April 16), https://variety.com/2018/digital/news/facebook-tipping-in-app-purchases-1202754904/.

Schmidt KM, Spann M, Zeithammer R (2015) Pay what you want as a marketing strategy in monopolistic and competitive markets. *Management Sci.* 61(6):1217–1236.

Shriver SK, Nair HS, Hofstetter R (2013) Social ties and user-generated content: Evidence from an online social network. *Management Sci.* 59(6):1425–1443.

Simonsohn U, Ariely D (2008) When rational sellers face nonrational buyers: Evidence from herding on eBay. *Management Sci.* 54(9):1624–1637.

Sjöblom M, Hamari J (2017) Why do people watch others play video games? An empirical study on the motivations of Twitch users. *Comput. Human Behav.* 75(10):985–996.

Sudhir K, Roy S, Cherian M (2016) Do sympathy biases induce charitable giving? The effects of advertising content. *Marketing Sci.* 35(6):849–869.

Sundie JM, Kenrick DT, Griskevicius V, Tybur JM, Vohs KD, Beal DJ (2011) Peacocks, Porsches, and Thorstein Veblen: Conspicuous consumption as a sexual signaling system. *J. Personality Soc. Psych.* 100(4):664–680.

Tang JC, Venolia G, Inkpen KM (2016) Meerkat and Periscope: I stream, you stream, apps stream for live streams. *Proc. 2016 CHI Conf. Human Factors Comput. Systems* (ACM, San Jose, CA), 4770–4780.

The Economist (2014) An itch to Twitch. Accessed May 25, 2021, https://www.economist.com/blogs/schumpeter/2014/05/live-video-streaming.

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

Tucker C, Zhang J (2011) How does popularity information affect choice? A field experiment. *Management Sci.* 57(5):828–842.

Zhang J (2010) The sound of silence: Observational learning in the US kidney market. *Marketing Sci.* 29(2):315–335.

Zhang J, Liu P (2012) Rational herding in microloan markets. *Management Sci.* 58(5):892–912.

Zhang X, Zhu F (2011) Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *Amer. Econom. Rev.* 101(4):1601–1615.