Information Systems Research

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

Fun Shopping: A Randomized Field Experiment on Gamification

Yi-Jen (Ian) Ho, Siyuan Liu, Lei Wang

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

INFORMATION SYSTEMS RESEARCH
Articles in Advance, pp. 1–19
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

Fun Shopping: A Randomized Field Experiment on Gamification

Yi-Jen (Ian) Ho,[a,*] Siyuan Liu,[a] Lei Wang[a]

[a] Smeal College of Business, Pennsylvania State University, University Park, Pennsylvania 16802
*Corresponding author

Contact: ian.ho@psu.edu, https://orcid.org/0000-0002-2480-480X (Y-J(I)H); siyuan@psu.edu, https://orcid.org/0000-0002-4315-8460 (SL); luw21@smeal.psu.edu, https://orcid.org/0000-0001-5531-9340 (LW)

Received: January 27, 2020
Revised: April 27, 2021; August 5, 2021; December 9, 2021
Accepted: December 20, 2021
Published Online in Articles in Advance: August 1, 2022

https://doi.org/10.1287/isre.2022.1147

Abstract. Gamification utilizes game-like features to engage participants, widely implemented in a variety of contexts. Such an IT-enabled engagement strategy serves as a marketing device to boost sales and customer loyalty. This study focuses on two significant game elements (i.e., badges and leaderboards) that promote consumer motivations and social comparisons. To qualify the impacts, we conduct a randomized field experiment at one of the largest shopping malls in Asia. In the experiment, we contrast the two elements against coupons regarding various shopping outcomes. A two-period design (consisting of the treatment and posttreatment periods) identifies the long-term behavior changes after the treatment removals. The main results suggest that badging and leaderboarding promote sales by 21.5% and 22.5% in the treatment period, respectively, whereas couponing delivers a more potent effect of 31.7%. In the posttreatment period, the gamification impacts remain significant compared with the baseline, but the influence of couponing fades out. Besides, the additional analyses document the salient heterogeneous treatment effects across demographics. We further discover the substantial differences in the within-group heterogeneity across the treatments. Specifically, badging is a balanced tool for attracting the general public, whereas leaderboarding is a double-edged sword that could encourage self-reinforcing or self-banishing. Finally, gamification brings more explorations that lead to additional sales and engagements. Overall, the robust results can be translated into actionable strategies to utilize gamification proactively.

History: Ravi Bapna, Senior Editor; Yili (Kevin) Hong, Associate Editor.

Keywords: gamification • badge • leaderboard • location-based technology • randomized field experiment • heterogeneous treatment effect

1. Introduction

Gamification, defined as using game-design elements in nongaming contexts, shows great potential across domains such as education, business, and health. The significance of gamification is highlighted by a $7.17 billion global market in 2019 and is projected to reach more than $40 billion by 2024 (ReportLink 2019). The prosperity of gamification comes from the growing need to enhance customer experiences and improve employee engagements. For instance, Google gamifies reimbursement processes to motivate employees to submit expense requests promptly. Deloitte exploits avatars and level-up systems to make senior executives more engaged in the training programs. Samsung launches a social loyalty game to cultivate relationships with various stakeholders (Patten 2016).

Gamification not only entertains, engages, and retains users, but spawns desired market outcomes. Such a useful engagement strategy starts drawing attention from researchers. The gamification literature generally focuses on the impacts of badges (Cavusoglu et al. 2015, Wang et al. 2020) and leaderboards (Bowey et al. 2015, Amo et al. 2018). Prior studies specifically establish positive relationships between gamification and the outcomes in education (Santhanam et al. 2016), business (Shao et al. 2019), and health (Wu et al. 2015, Hydari et al. 2019), etc.

Despite the growing evidence advocating the effectiveness of gamification, it remains unclear what underlying mechanisms drive users to behave as expected (Liu et al. 2017). The knowledge in marketing is particularly underdeveloped. Though game elements are used to enhance shopping experiences, it is challenging for researchers to provide rigorous causal inferences using fine granular data (Xi and Hamari 2020). Beyond the proof of causality, it is critical in the marketing theme to contrast gamification, a nonmonetary motivation, with couponing, a monetary incentive. Besides, though marketers adopt gamification to develop long-term behavior changes, temporal analyses are generally ignored in the literature (Tobon et al. 2020). This reminds us of the need to investigate the long-term impacts on desired outcomes. Finally, researchers have not proactively looked at how user characteristics affect gamification-driven behaviors (Hofacker et al. 2016). Identifying the heterogeneous effects not only theoretically illustrates how specific segments would respond to IT-enabled gamification but practically showcases how such an IT artifact could be utilized to target consumers.

The aforementioned research gaps jointly lead to a need for rigorous investigations on the impacts of gamification on consumer engagements. Specifically, we ask the following research questions: (1) Would gamification (i.e., badges and leaderboards) help promote consumer shopping engagements? (2) If so, how would gamification differ from coupons? (3) Would gamification be a sustainable strategy in the long run after games stop? (4) How would heterogeneous consumers respond to various game features?

These questions are fundamental because gamification broadly helps us understand how human beings could be motivated. In the gamification literature, self-determination theory (SDT) and social comparisons theory (SCT) serve as theoretical foundations (Tobon et al. 2020). In the SDT framework, gamification serves as *intrinsic* and *extrinsic* interventions through various game elements that support competence and autonomy (Peng et al. 2012, Domínguez et al. 2013). Social comparisons are adopted to theorize the impacts of competing games, such as leaderboards (Wu et al. 2015, Hydari et al. 2019). In the information systems (IS) research, the two theories are also employed to explain the effects of gamification in education (Biles et al. 2018), e-commerce (Shao et al. 2019), user-generated content (UGC) (Cavusoglu et al. 2015, Wang et al. 2020), and health (Wu et al. 2015, Hydari et al. 2019). We carefully follow the literature to construct the theoretical supports in the literature review.

In this study, we attempt to simultaneously stimulate self-determination, social comparisons, and monetary incentives using collectible badges, public leaderboards, and coupons. We briefly discuss the rationales behind the alignments between stimuli and mechanisms. First, we make badges without monetary value and are privately accessible only. The badge-collecting behavior should be driven by self-satisfaction, activating intrinsic and extrinsic motivations. Second, a leaderboard publicly ranks each individual's performance of interest (e.g., shopping distance). Such visibility of performance allows a focal consumer to compare the consumer's self with peers socially. Finally, couponing offers additional monetary rewards to entice consumers to engage. In short, in a unified research setting, the three treatments jointly demonstrate the similarities and differences across the three motivations.

To answer our research questions, we collaborate with one of the largest shopping malls in Asia. We conduct large-scale randomized field experiments with a two-period design, qualifying the impact of gamification on off-line shopping engagements powered by location-based technologies. A ubiquitous Wi-Fi system is exploited to identify each visitor, track the visitor's shopping trajectories, and deliver the treatments of interest.[1] We randomly assign visitors to the four experimental groups, including *Badge*, *Leaderboard*, *Coupon*, and *Control*. Group *Badge* is given a unique badge designed for a specific store category when making associated purchases.[2] Second, group *Leaderboard* earns points for shopping distances and store visits to promote ranks on a public leaderboard. Third, group *Coupon* is financially incentivized by a 20%-off coupon for a random store. Finally, group *Control* receives no incentive but a greeting message. These exogenous variations enable us to quantify each treatment's effect and compare the magnitudes across them (List and Rasul 2011). Moreover, our two-period design quantifies the treatment vitalities by considering a *treatment period* followed by a *posttreatment period*. Specifically, we deliver the treatments in the first period and withdraw them in the second one.

The empirical analyses generate two sets of findings regarding shopping expenditures, distances, and store visits. First, badges, leaderboards, and coupons effectively incentivize consumer shopping engagements. Game elements show their potent effects on nonsales actions, whereas couponing demonstrates a dominant impact on spending. Quantitatively, badging and leaderboarding generate 21.5% and 22.5% more revenues in the treatment period, and coupons bring a 31.7% increase in sales.[3] In contrast, group *Badge* (*Leaderboard*) is encouraged to stroll 25% (28.4%) farther and visit 37.9% (33.4%) more stores, reflecting an additional 9.4% (12.8%) and 27.1% (22.6%) of activities on top of the couponing effect. Moreover, gamification-based engagements stay significant even after games stop, but coupons' impact fades out sharply. Unlike couponing, gamification reshapes consumer behaviors and continues the engaged shopping. Such behavioral changes demonstrate the long-term influence of gamification, partially attributed to promoting nonsales activities in the treatment period.

Second, consumer heterogeneities play a critical role in responding to the treatments. Women prefer coupons, whereas men are more into games. Also, gamification is more attractive to youths, but couponing seems a one-for-all strategy. Regarding the moderating effect of income, we find that badges (leaderboards) are more compelling to the high-income (low-income) segment. We also reveal how the consumer pool reacts to the treatments differently using the within-group analyses. Group *Leaderboard* has the highest within-group heterogeneity, followed by groups *Badge* and *Coupon*. This suggests that social comparisons on leaderboards could be a double-edged sword, wherein more (less) engaging subjects may develop a self-reinforcing (self-banishing) process. In contrast, badging with moderate heterogeneity is a balanced method to engage the general public.

Besides, the extended analyses examine potential mechanisms behind the salient gamification effects. Specifically, incentivized explorations are considered. The empirics show the exploration effect in the treatment period that becomes marginal after the removals of game elements. Such a *temporal* effect may be due to the "call for action" nature, whereas the long-lasting behavior changes come from consumers' inner motivations. We finally rule out the alternative explanations (i.e., companion and seasonality) using subsamples. The robust results confirm the main findings when considering noncompanioning shopping cases and nonstudent consumers (above the age of 22).

Our research makes a series of contributions to the gamification literature. First, we complement the literature by exploring consumer behaviors in a new context (i.e., a brick-and-mortar mall). Specifically, construct a novel *online-to–off-line* gamified environment using location-based technologies. We also extend the scope of prior studies while studying multiple game elements and monetary incentives simultaneously. Second, to respond to the call for causal inferences (Hamari et al. 2014), we conduct randomized field experiments. Thousands of general shoppers are randomly assigned to stimuli in a well-controlled experiment to obtain empirical regularities. Third, the gamification literature still lacks analyses with a broader time horizon (Tobon et al. 2020). Our study considers both treatment and posttreatment periods to capture potential behavior changes after gamification is removed. Finally, we comprehensively explore the heterogeneous gamification effects with an eye on the moderation roles of demographics. We also characterize the distinct natures of motivation mechanisms by comparing the within-group heterogeneities across gamifying and couponing. These new, robust empirics support future theory developments.

The findings can be translated into several actionable recommendations. For the management of a shopping center, gamification can be utilized to promote the outcomes of interest by cultivating loyal customer relationships. When participating in location-based games, consumers develop attachments to the mall through more exposure to new excitements, which lead to better shopping experiences. Besides, the discovered heterogeneous treatment effect (HTEs) enrich targeting strategies using gamification. Moreover, this study has guidance for the gamification design. A designer exploits IT-enabled games that are readily configured in a cost-efficient way. For example, traffic can be redirected to specific stores by providing attractive game elements (e.g., eyeball-catching badges). Finally, by tracking locations and profiling customers, designers have the opportunity to construct real-time game features based on customers' trajectories and preferences.

The rest of the paper proceeds as follows. We first develop the theoretical foundation by reviewing the literature in Section 2. We detail the experimentation in Section 3. Section 4 discusses the empirical analyses, including models, results, and economic significances. The mechanism is investigated in Section 5. We conclude this research with both theoretical contributions and managerial implications in Section 6.

2. Related Literature

We start reviewing prior work on gamification in the IS literature as a whole. These studies help us understand how gamification affects user engagements in various contexts. Then, we focus on two specific elements, badges and leaderboards, and discuss the related theoretical background. The overall conceptual research framework is summarized at the end.

2.1. Gamification

Gamification utilizes game elements to engage users in nongaming contexts (Deterding et al. 2011).[4] Badges and leaderboards are the most significant two because of their wide implementations (Bittner and Shipper 2014, Sigala 2015). To understand the emergence of badging and leaderboarding, the literature mainly focuses on several themes, including user-generated content, health, and education. In the UGC domain, Cavusoglu et al. (2015) document the efficacy of a badge system on stimulating content-generating at Stack Overflow. Wang and Sanders (2019) similarly show that badging leads to more and longer reviews posted. In the health context, a few studies examine how leaderboards motivate health activities. Benefiting from the embedded leaderboards on Nike+ and Fitbit wearables, Wu et al. (2015) and Hydari et al. (2019) report students' promoted physical activities. As to learning settings, prior studies identify the positive relationship among badges, learning motivations, and test performances (Abramovich et al. 2013, Denny et al. 2018).

The literature starts examining the impact of gamification on profitability, yet the related evidence remains limited. Only a few focus on how gamification entices consumers to be active at e-commerce. Hamari (2017) finds that the badge feature encourages college students to carry out more transactions, and Shao et al. (2019) show that badging drives impulsive purchases. Given the online nature, it is technically challenging to examine the relationship between gamification and off-line shopping in reality. In this regard, this paper complements the literature by exploring consumer behaviors in a brick-and-mortar mall. Specifically, we construct a novel online-to–off-line gamified environment. Table 1 contrast this study with the representative reference.

Our research broadly examines gamification impacts on user engagements, taking a wider lens than prior literature in several dimensions. We first inclusively

Table 1. Selected Literature of Gamification

		Effect of interest			Research design		Analysis	
Reference	Context	Badge	Leaderboard	Monetary	Methodology	Samples	Posttreatment	Heterogeneity
This study	Off-line shopping	√	√	√	Randomized field experiment	Mall visitors	√	Demographics and cross-stimuli
Shao et al. (2019)	Ecommerce	√		√	Survey	Platform users		
Hamari (2017)	Local college community	√			Archival analysis	College students		
Cavusoglu et al. (2015)	User-generated content	√			Archival analysis	Platform users		
Wang and Sanders (2019)	User-generated content	√		√	Laboratory experiment	Amazon MTurk		
Hydari et al. (2019)	Health		√		Archival analysis and survey	College students		Prior performances
Wu et al. (2015)	Health		√		Archival analysis and survey	College students		
Abramovich et al. (2013)	Education	√			Laboratory experiment and survey	Middle-school students		Prior knowledge
Denny et al. (2018)	Education	√			Randomized field experiment	College students		

compare multiple engagement strategies. Prior studies mainly focus on either badging (Hamari 2017) or leaderboarding (Hydari et al. 2019)—one or the other in isolation. In contrast, we study both game elements in a single context. While considering distinct motivation mechanisms, we further benchmark gamification against monetary incentives commonly adopted in practice but generally ignored in the related literature. Second, this study qualifies the effects in a randomized field experiment. Hamari et al. (2014) identify the need for rigorous causal inferences. The existing literature heavily relies on surveys, laboratory experiments, and archival analyses using specific subject groups (e.g., college students). Complementarily, we randomly assign stimuli to thousands of general shoppers in a well-controlled environment to obtain empirical regularities.

Third, the literature still lacks the analysis with a broader time horizon (Tobon et al. 2020). Minimal attention is paid to longitudinal analyses (Hamari et al. 2014). We specifically respond to the call by looking at potential behavior changes after gamification removals. In this regard, our experiment is novel and designed to incorporate both treatment and posttreatment periods to capture such temporal changes. Finally, we explore the heterogeneous effects. Without an eye on characteristic-driven moderations, prior studies mainly either treat demographics as controls (e.g., Shao et al. 2019) or use them in balance checks (Shang and Lin 2013, Wang and Sanders 2019). Abramovich et al. (2013) take an initial step toward study the moderating role of prior knowledge in an e-learning system. Hydari et al. (2019) then consider how previous exercise habits affect later performances on leaderboards. To shed light on how consumer heterogeneity interacts with off-line gamified shopping environments, we consider the moderation of demographics (i.e., gender, age, and income) and the extent of within-group heterogeneities across gamifying and couponing.

2.2. Badge and Self-Determination

Badges, as digital representations of knowledge and experiences, encourage, recognize, and acknowledge user achievements across digital platforms (Kwon et al. 2015). The gamification literature starts studying badges in various settings with mixed findings, including positive, negative, and no effects. Cavusoglu et al. (2015) find a positive relationship between the number of collected badges and the motivated willingness to contribute at Stack Overflow. Yet Higashi (2012) shows the opposite, wherein the more badges earned, the worse performances are made in an online learning system. Studying a college community, Hamari (2013) also shows that badging does not guarantee an increase in user engagements. In short, despite the growing interest in digital badges, the literature still lacks consistent findings and new evidence in off-line settings, inspiring us to investigate the impact of gamification on off-line business activities.

To address these mixed findings and explore the motivation mechanism behind them, scholars generally adopt self-determination theory as the theoretical foundation (Tobon et al. 2020). SDT defines a motivation framework to theorize how humans initiate and regulate behaviors (Deci and Ryan 1985). The theory suggests that people can be motivated via interventions and become self-determining when their needs for *competence*, *autonomy*, and *relatedness* are fulfilled (Ryan and Deci 2000). SDT further classifies motivations into intrinsic and extrinsic ones. Intrinsic drivers refer to three basic psychological needs, whereas extrinsic interventions refer to rewards and punishments externally imposed. It is also worth noting that extrinsic interventions can be internalized to fulfill the three intrinsic needs (Ryan and Deci 2000). Given the blurry boundary between two types of motivations, the gamification literature holds different views on the motivation, including intrinsic motivations (Cavusoglu et al. 2015, Picone et al. 2019), extrinsic motivations (Hanus and Fox 2015), and a mixture of the two (Hassan 2016, Biles et al. 2018).

We closely follow the literature and root our work in the self-determination theory framework. The literature states that badges motivate competence, autonomy, and relatedness. First, collected badges provide competence-relevant feedback, enhancing the competence from task accomplishments (Sailer et al. 2013). Badges can serve as system-certificated statuses for competence without user bragging (Antin and Churchill 2011). Second, a badge system usually offers various badges for different tasks and allows users autonomously to choose. Such autonomy enhancement is proved to be effective as Randall et al. (2013) discuss in self-regulated learning. Finally, badges promote relatedness to the environment (Hamari 2013). When earning a badge, a user naturally raises more inner attachments to the surroundings. Besides, SDT explains that external interventions can be transformed into intrinsic motivations. Such internalization occurs when an extrinsic intervention fulfills the three needs (Ryan and Deci 2000). We believe that badging helps internalize the external icons by recognizing a user's competent performance, allowing the user to have autonomous participation and enhancing the user's relatedness to the environment (Cavusoglu et al. 2015). Thus, though seen as extrinsic interventions, badges are generally internalized to nourish intrinsic motivations (Picone et al. 2019). We acknowledge that badges are a unique mixture of both intrinsic and extrinsic interventions.

We respond to the call for more empirical evidence in off-line settings (Sailer et al. 2013). We study an underexplored area and examine the impact of badges on off-line shopping engagements. In this study, digital badges are delivered as a clean treatment without social interactions (i.e., privately accessed badges) or monetary incentives. By doing so, we align badges with the self-determination motivations (Hofacker et al. 2016).

2.3. Leaderboard and Social Comparison

A leaderboard descendingly ranks players regarding their performances for a specific task (Duggan and Shoup 2013). Because of technological advancements, leaderboards are digitalized and accessed whenever needed, engaging users through social comparisons. Unlike physical leaderboards subjected to accessibility (e.g., a chart in the classroom), digital ones are much more prevalent (Hanus and Fox 2015). Leaderboarding is widely studied in education (Hanus and Fox 2015, Amo et al. 2018, Kwon and Özpolat 2020), management (Costa et al. 2013, Mollick and Rothbard 2014), and health (Wu et al. 2015, Hydari et al. 2019). However, prior studies report mixed findings of leaderboards' effectiveness. Though there exist positive correlations between leaderboards and user engagements (Huang and Hew 2015, Wu et al. 2015, Zhang et al. 2021), other studies document negative (Hanus and Fox 2015) or no impacts (Costa et al. 2013).

Despite mixed empirics, prior literature generally adopts social comparison theory to theorize the influence of leaderboards (Wu et al. 2015, Hydari et al. 2019). Individuals perform self-assessments by comparing themselves with others (Festinger 1954). These comparisons include *upward* and *downward* comparisons (Buunk and Gibbons 2007). By displaying players' ranks, a leaderboard enables players to perform both types of comparisons at the same time (Costa et al. 2013, Christy and Fox 2014). As Dijkstra et al. (2008) argue, upward comparisons evoke self-inferiority leading to a negative impact (i.e., desperation), whereas downward comparisons generate self-superiority resulting in a positive effect. Yet self-inferiority can be turned into proactive determination to improve state quotes (Fotaris et al. 2016), and self-superiority can also bring extra pressures to maintain the current status (Wells and Skowronski 2012). In short, the literature still leaves the effects of leaderboards unclear from the social comparison perspective.

We thus reexamine the impacts of leaderboards given the unanswered question. Considering the lack of direct evidence, we conduct a randomized field experiment to study the causality that complements survey- and laboratory-based empirics (Wu et al. 2015, Wang et al. 2020). As a double-edged sword, a leaderboard is perceived to either evoke participants' egos or confine their autonomy (Amo et al. 2018). Hydari et al. (2019) argue that such differences originate from individual heterogeneity. Thus, these mixed effects might be explained through a high-resolution lens. In this regard, we decompose the overall impact into each individual's response while considering the moderations of demographics. This study is expected to contribute to the literature by comprehensively characterizing the nature of leaderboarding.

2.4. Research Framework

W attempt to empirically quantify the effectiveness of badges and leaderboards. As prior studies back up the mechanisms using SDT (Deci and Ryan 1985) and SCT (Festinger 1954), we contrast the two nonmonetary game elements with monetary incentives (i.e., coupons). Badges, leaderboards, and coupons serve as the building blocks in the research framework illustrated in Figure 1. To fulfill the needs of causality, we choose a shopping mall as the playfield to experiment with gamification and couponing. The results provide the causal evidence to solve the aforementioned literature discrepancies of gamification.

Besides, prior gamification literature has not paid much attention to temporal analyses, making it challenging to predict the long-term effects. In this regard, we discuss how badging potentially results in long-lasting impacts by stretching self-determination motivations. The SDT-based long-term effects are recognized (Williams et al. 2009, Teixeira et al. 2012). Specifically, perceived competence and autonomy support long-term tobacco abstinence (Williams et al. 2009) and weight control (Teixeira et al. 2012). As discussed in Section 2.2, badges, a mixture of intrinsic and extrinsic motivations, are useful to motivate competence, autonomy, and relatedness. We conjecture that group *Badge* will demonstrate long-term engagements. In addition, we discuss the potential posttreatment effect for leaderboards by extending SCT. Social comparisons, especially upward comparisons, show positive impacts on students' academic performances in the long run (Wehrens et al. 2010). Witnessing peers' successes evokes one's sense of personal honor immediately, which nourishes self-confidence and competence thereafter (Buunk et al. 1990). As prior literature states that social comparisons can be lasting motivations, we expect that the subjects in group *Leaderboard* will continue being engaged.

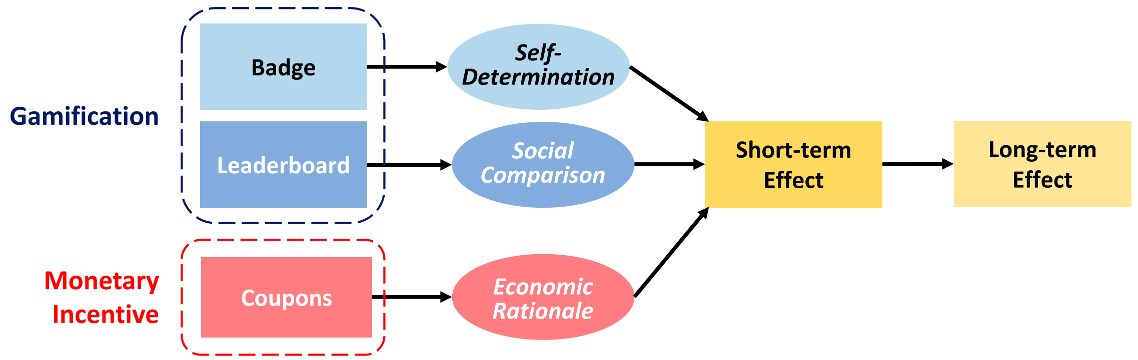
3. Experimentation

Based on the theoretical framework in Section 2, we investigate how gamification features affect consumer shopping engagements in an off-line setting. We conduct a randomized field experiment, partnering with one of the largest shopping malls in Asia. To avoid conscious bias, we follow the design of List and Rasul (2011), wherein the subjects (i.e., mall visitors) are completely unaware of being observed. We then observe the subjects' responses to the treatments in a controlled environment to isolate the treatment effects from other confounders (List and Rasul 2011). The shopping center is located in an Asian major city and hosts more than 300 international and domestic brands and a variety of amenities and restaurants.[5] On average, it attracts around 150,000 visitors daily and has a revenue of $450 million annually. The shopping mall offers visitors free Wi-Fi and has an official shopping portal that posts the latest news. Given the valuable information (e.g., events and promotions) and convenience, around 90% of visitors take advantage of the Wi-Fi internet and the portal to enhance their shopping experiences.[6] Simple login (using either WeChat or a member account) is required to access the Wi-Fi service for the first time only, and the service automatically connects thereafter. We utilize the ubiquitous Wi-Fi system to identify each visitor, track the user's shopping trajectories via indoor Wi-Fi trackers, and deliver the treatments of interest (i.e., game features and coupons).

3.1. Experimentation Design

In the experiment, we attempt to (1) identify the impact of gamification on consumer shopping engagements, (2) compare gamification features and couponing, and (3) measure the posttreatment effects of the various treatments. We need a two-stage experiment wherein multiple treatments are given. We start discussing

Figure 1. (Color online) Research Framework



the experimentation with the timeline. The experiment lasts for eight weeks, from July 1 to August 25, 2018.

As Figure 2 illustrates, we experiment with a two-period setup while considering both treatment and posttreatment periods.

Accordingly, the entire eight-week window is divided into two periods. We define the first four weeks (from July 1 to 28) and the second four weeks (from July 29 to August 25) as the treatment and posttreatment periods. We complete the random assignment process on the first day of the treatment period, starting at 10 a.m. and ending at 10 p.m. In other words, the subjects are randomly selected from the visitor pool on July 1 only.[7] When a visitor arrives whose device is connected to the Wi-Fi service, we randomly assign the visitor to either one of the treatment groups or the control group. Once the assignment is determined, a notification is immediately pushed to the visitor's device through WeChat and the shopping portal. The corresponding instruction and hyperlink for the group are embedded in the message to browse the program page. The subject is automatically enrolled in the assigned treatment program, if any, and takes no further action to participate.[8] The treatment reminders are sent whenever the subjects revisit the mall. Though we observe the subjects' shopping responses to the treatments through the entire eight-week window, the treatments are given only during the first four weeks. We immediately stop all the treatments on the first day of the posttreatment period, July 29. To avoid potential strategic responses, we do not notify the subjects in advance when the treatments end. The control group remains not receiving any treatment during the entire experiment.

We experiment with badging and leaderboarding because of their significance as discussed earlier. For the badge treatment, a unique badge is designed for each store category.[9] A visitor in group *Badge* is given badges for making purchases in the corresponding categories. The visitor can check the visitor's collection page any time that can be assessed by the visitor's self only. To avoid potential social interactions, we restrict the focal visitor as the only person who can access the visitor's badge collection.[10] For the leaderboard feature, visitors can earn points for the distance they walk and the number of store visits they make.[11] The subjects in group *Leaderboard* accumulate points to promote their ranks on the leaderboard that refreshes every minute.[12] In this regard, a consumer may compete with other leaderboard participants for a higher rank. Besides, we incorporate a coupon treatment into the experiment to compare nonmonetary incentives (i.e., gamification features) with coupon discounts. We randomly select stores across the nine categories and craft corresponding 20%-off coupons to avoid the store-selection bias. A subject in group *Coupon* is given a random digital coupon (i.e., a QR code) good for one-time use in a specific store. Once the coupon is redeemed, the subject does not receive a new one for subsequent visits. The treatments are summarized in Table 2.

Figure 2. (Color online) Timeline of Experiment

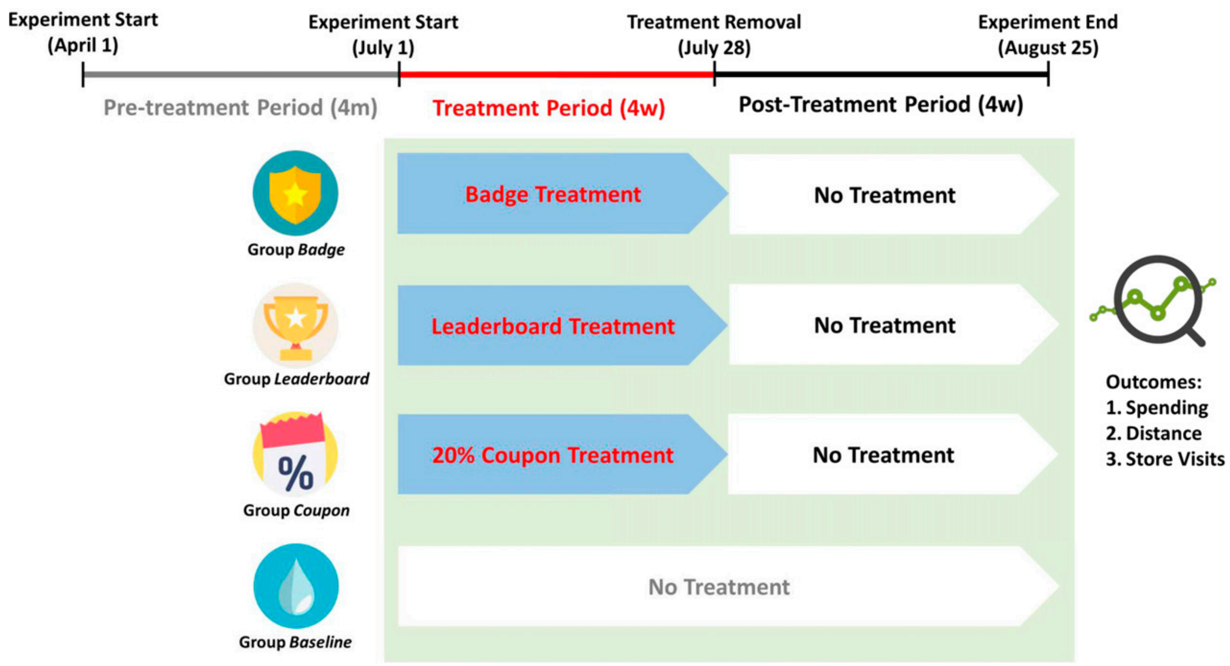


Table 2. Experimental Groups

Group	Treatment	*N*
Badge	A visitor collects a unique badge from the nine categorical badges for the visitor's first purchase of the category.	2,000
Leaderboard	A visitor earns one point for every 0.06-mile walk (100 m) and 20 points for every store visit.	2,000
Coupon	A visitor is randomly offered a 20%-off coupon (one-time use) for a store in one of the nine categories.	2,000
Control	—	2,000

Figure 3 illustrates the user interfaces of each group. In panel (a), a participant in group *Badge* has already collected four different badges, meaning that the participant has shopped in four corresponding categories. In panel (b), the focal subject, as an active member in group *Leaderboard*, has accumulated 8,230 points. The subject ranks second in the group at this moment. Panel (c) illustrates how a coupon is delivered to a subject: the instruction details the discount rate and the store location. Besides these three treatment groups, we have the control group receiving a simple greeting, shown in panel (d).

3.2. Balance Check and Descriptive Statistic

It is the prerequisite for the following analyses having comparability across groups. Our validation strategy checks both preassignment and postassignment balances. For the preassignment check, we start recording visitors' behaviors two months prior to the experiment (see Figure 2). We profile each visitor regarding spending, shopping distance, and store visits based on the visitor's shopping trajectories. These individual profiles are used to examine whether the groups have similar shopping patterns before the random assignments. For the postassignment check, we conduct a short survey right after a visitor is assigned to a group.[13] The survey collects a set of demographics consisting of gender, age, and income. The initial response rate is 78.4%. For nonresponders, a friendly reminder is sent for every mall visit. Also, sales representatives collect the satisfaction survey during the checkout process, asking for the same demographics and satisfaction feedback. These two jointly recover the demographics of 96.1% of the pool. We validate the balances using pairwise *t*-tests and joint *F*-tests in Table 3. Panel A reports the results for the preassignment checks. The insignificances show that the composition of each group is statistically identical before the assignments. A series of insignificant results suggest indifference across the groups after the assignments in panel B.

To measure the impacts of gamification and couponing on shopping behaviors, we consider three outcomes of each subject, including shopping amounts, shopping distances, and the number of store visits. To capture the shopping transactions, we collaborate with sales representatives to match sales records and visitor identifications. In every checkout process, representatives ensure that each visitor's identification is matched with the visitor's device identification.[14] In-door Wi-Fi trackers track shopping trajectories used to operationalize variables, such as distances and store visits (i.e., stops).[15] We construct the fine granular data at the individual–period level. Table 4 reports the descriptive statistics of the data. The three outcomes, $Money_{ijt}$, $Distance_{ijt}$, and $StoreVisits_{ijt}$, refer to the amount spent, the walking distance, and the number of store visits by visitor i in group j at period t. $Money_{ijt}$ has a mean of 140.18, indicating that a visitor on average spends \$140.18 in a four-week period. The average shopping distance is 2.35 miles with a high standard deviation of 1.06. Also, visitors typically visit 51 stores in the mall. Groups *Badge*, *Leaderboard*, and *Coupon* are binary indicators for the group assignments,

Figure 3. (Color online) User Interface of Treatment

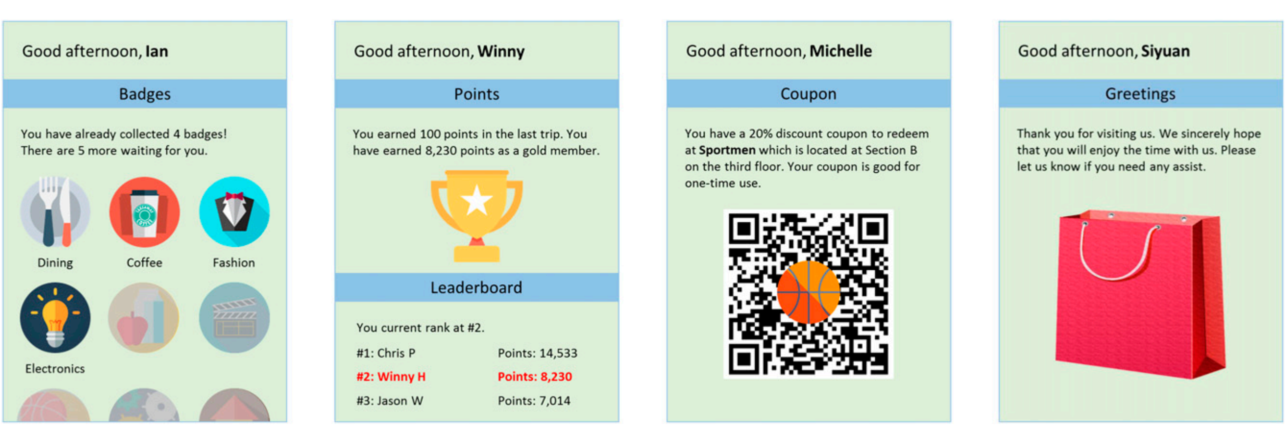


Table 3. Balances

Panel A. Preassignment checks			
Pairwise *t*-tests			
	Badge versus *Leaderboard*	*Badge* versus *Coupon*	*Badge* versus *Control*
Money	0.251	1.090	−0.235
Distance	1.463	0.160	1.195
Store_Visits	0.475	0.362	0.080
	Leaderboard versus *Coupon*	*Leaderboard* versus *Control*	*Coupon* versus *Control*
Money	0.845	−0.486	−1.319
Distance	−1.306	−0.295	1.034
Store_Visits	−0.105	−0.398	−0.287
Joint *F*-tests			
	Money	*Distance*	*Store_Visits*
F-statistics	0.903	1.081	0.102

Panel B. Postassignment checks Pairwise *t*-tests			
	Badge versus *Leaderboard*	*Badge* versus *Coupon*	*Badge* versus *Control*
Gender	−0.396	−0.477	0.290
Age	−0.354	0.518	0.351
Income	0.067	1.126	0.188
iOS	1.515	0.154	−0.043
	Leaderboard versus *Coupon*	*Leaderboard* versus *Control*	*Coupon* versus *Control*
Gender	−0.590	−0.918	−0.281
Age	0.759	0.645	−0.193
Income	1.278	0.234	−1.464
iOS	−1.361	−1.554	−1.190

Joint *F*-tests				
	Gender	*Age*	*Income*	*iOS*
F-statistics	0.369	0.517	0.841	0.560

Notes. $N = 8{,}000$. All *t*- and *F*-statistics are statistically insignificant.

and each group accounts for 25.0% of the total observations. The composition of gender is representative, and 37.9% of the subject pool are male subjects (*Gender* = 1). Age ranges from 16 to 60, and the average age is 33.9. The visitors have an average annual income of \$21,000 with a high standard deviation of \$22,119. As to device types, 41.6% (58.4%) of the subjects have devices running iOS (i.e., Android or other operating systems).[16]

3.3. Model-free Evidence

A series of Tukey's tests are conducted as the model-free evidence. Table 5 details the results. We start analyzing the logarithm values of the dependent variables in the treatment period. We prefer Tukey's tests because they compare means across groups simultaneously, whereas a *t*-test focuses on pairwise comparisons. In panel A, three treatment groups are more engaged than the control. Two game features and coupons effectively incentivize the visitors to shop more, stroll farther, and stop by more stores. Among the treatment groups, the effect of couponing on spending is more potent than gamification. In contrast, the badge and leaderboard features show the superiority in the shopping distances and the number of store visits. If we take a closer look, group *Leaderboard* is willing to walk farther. Both badge and leaderboard groups have no statistical differences regarding the number of store visits.

Tukey's tests are also applied to the subjects' behaviors in the posttreatment period. The results are

Table 4. Descriptive Statistics

	Mean	Standard deviation	Minimum	Maximum
Money	140.180	414.4	53.00	2,455
Distance	2.345	1.057	0.705	7.578
Store_Visits	51.14	18.11	19	157
Badge	0.250	0.433	0	1
Leaderboard	0.250	0.433	0	1
Coupon	0.250	0.433	0	1
Post	0.500	0.500	0	1
Gender	0.379	0.480	0	1
Age	33.91	9.562	16	60
Income	21,004	22,119	500	60,000
iOS	0.416	0.484	0	1
N		8,000		

Note. The outcomes are aggregated at the individual–period level.

Table 5. Tukey's Tests

	Badge	Leaderboard	Coupon	Control
Panel A. Mean differences in the treatment period				
Money				
Badge	—			
Leaderboard	−0.063	—		
Coupon	−0.242***	−0.240***	—	
Control	0.169***	0.181***	0.409***	—
Distance				
Badge	—			
Leaderboard	−0.061	—		
Coupon	0.100***	0.131***	—	
Control	0.278***	0.305***	0.172***	—
Store visits				
Badge	—			
Leaderboard	0.056	—		
Coupon	0.262***	0.221***	—	
Control	0.319***	0.248***	0.095***	—
Panel B. Mean differences in the posttreatment period				
Money				
Badge	—			
Leaderboard	0.073***	—		
Coupon	0.125***	0.068***	—	
Control	0.127***	0.104***	0.023	—
Distance				
Badge	—			
Leaderboard	0.037	—		
Coupon	0.151***	0.136***	—	
Control	0.173***	0.146***	0.065	—
Store visits				
Badge	—			
Leaderboard	0.094***	—		
Coupon	0.209***	0.126***	—	
Control	0.218***	0.154***	0.010	—

Note. $N = 8{,}000$.
*** 0.01 level of significance.

summarized in panel B. Interestingly, the mean differences between the gamification and control groups remain statistically significant, suggesting that the gamification-based engagements stay notable even after the games stop. In contrast, the couponing effect fades out after coupons expire. In other words, the visitors tend to maintain shopping habits specifically reshaped by gamification. These findings are much different from those found in the treatment period, wherein couponing is the most effective strategy. To quantify the treatment effects, we are motivated to conduct empirical analyses in Section 4.

4. Empirical Analysis

In this section, we start estimating the average treatment effects and highlight the corresponding economic significances. Then, we investigate the heterogeneous treatment effects of demographic moderators across the groups. Finally, we make a closer examination of the heterogeneity within each experimental group and characterize the different natures of gamification and couponing.

4.1. Average Treatment Effect (ATE)

Tukey's tests provide an overview of the treatment effects, and multivariate regression analyses help distill the ATEs while controlling other covariates. We specify a linear regression model as

$$\begin{aligned} outcome_{ijt}^{k} = {} & \beta_0 + \beta_1 Badge_{ij} + \beta_2 Leaderboard_{ij} + \beta_3 Coupon_{ij} \\ & + \beta_4 Post_t + \beta_5 Badge_j * Post_{it} + \beta_6 Leaderboard_{ij} \\ & * Post_t + \beta_7 Coupon_{ij} * Post_t + \alpha_1 Gender_i \\ & + \alpha_2 age_i + \alpha_3 income_i + \alpha_4 iOS_i + \varepsilon_{ijt}, \end{aligned} \tag{1}$$

where superscript k (= 1, 2, and 3) denotes the logarithm value of the outcome ($money_{ijt}$, $distance_{ijt}$, and $visit_{ijt}$) by visitor i in group j at time t. Variable $Post_t$ is a binary dummy indicating the treatment period ($Post_t = 0$) and the posttreatment period ($Post_t = 1$). The main effects capture the treatment effects. The model also includes the two-way interactions that show the experimental groups' behavioral changes between the treatment and the posttreatment periods. Finally, demographics serve as additional controls, including gender, age, income, and operating systems.[17] Ordinary least squares are used to estimate the regression with the clustered errors at the individual level.

Table 6 summarizes the results for the three outcomes. For the shopping amounts, the treatment effects are positive and statistically significant. In other words, badges, leaderboards, and coupons motivate the treated visitors to spend more than those untreated. The coefficient of $Post_t$ is insignificant, suggesting that the control group retains time-invariant shopping patterns across the entire experiment.[18] $Badge_{ij} * Post_t$ and $Leaderboard_{ij} * Post_t$ have negative, significant coefficients. The gamification effects decline in the posttreatment period, implying that the gamification groups spend less after the game features stop. Similarly, the coupon group's purchase power dramatically decreases because of the lack of monetary incentives. Columns (2) and (3) show the analogous evidence regarding the shopping distance and the number of store visits, respectively. The main effects are statistically significant, so the three treatments positively influence distance and store visits in the first four weeks. Groups *Badge* and *Leaderboard* remain engaged in the second period, but group *Coupon* relatively loses interest in exploring. As to the control variables, females generally enjoy shopping more than males. Young generations are more energetic but less willing to spend money in the mall. People with high income typically have high purchasing power, but device types do not affect the outcomes.

We quantify the economic significance of the treatment effects using the estimates in Table 6. Because the dependent variables are the logarithm values, we exponentiate a coefficient and interpret it as percentage increases responding to a one-unit increase in the covariate of interest. In the treatment period, collecting badges and competing on leaderboards inspire the subjects to spend 21.5% (i.e., $exp(0.195) - 1$) and 22.5% more than the baseline subjects, respectively. Couponing is effective, boosting spending by 31.7%. Given the control group's average spending of \$125.15, these increases suggest that the two game elements and couponing generate \$26.91, \$28.16, and \$39.67 more in revenues (per person), respectively. In the posttreatment period, groups *Badge* and *Leaderboard* continue bringing additional sales of \$11.64 (i.e., $125.15 \times (exp(0.195 - 0.106) - 1)$) and \$8.54, respectively, compared with the control group. In contrast, the coupon group behaves nearly the same as the baseline, contributing an additional 25 cents that is neglectable. In general, women are willing to spend 22.4% more than men, all else equal. A 10% increase in age and income result in 5.1% (i.e., $1 - 1.10^{0.523}$) and 7.5% increases in spending. We now turn to discuss the other two outcomes.

Table 6. Average Treatment Effects

	money	*distance*	*store visits*
Intercept	4.727 (0.038)***	0.502 (0.021)***	−3.535 (0.023)***
Badge	0.195 (0.002)***	0.223 (0.002)***	−0.321 (0.002)***
Leaderboard	0.203 (0.002)***	0.250 (0.002)***	−0.288 (0.002)***
Coupon	0.275 (0.002)***	0.145 (0.002)***	−0.103 (0.002)***
Post	0.823 (0.995)	1.106 (0.994)	1.065 (1.000)
Badge ∗ Post	−0.106 (0.003)***	−0.043 (0.003)***	−0.070 (0.003)***
Leaderboard ∗ Post	−0.137 (0.003)***	−0.091 (0.003)***	−0.106 (0.003)***
Coupon ∗ Post	−0.273 (0.003)***	−0.139 (0.003)***	−0.105 (0.003)***
Gender	−0.254 (0.020)***	−0.165 (0.021)***	−0.239 (0.021)***
age	0.523 (0.116)***	−0.626 (0.115)***	−0.814 (0.113)***
income	0.763 (0.122)***	0.319 (0.124)***	0.156 (0.122)
iOS	0.036 (0.021)	0.030 (0.020)	0.028 (0.021)
Adjusted R^2	0.646	0.425	0.429
N		16,000	

*** 0.01 and 0.05 levels of significance.

For shopping distance, groups *Badge*, *Leaderboard*, and *Coupon* walk 25%, 28.4%, and 15.6% farther in the first period, resulting in an increase of 0.463, 0.525, and 0.289 miles, respectively.[19] In the second period, groups *Badge* and *Leaderboard* continue strolling extra mileages of 0.365 and 0.319 miles. Also, the visitors who collect badges (compete on the leaderboard) tend to check out 37.9% (33.4%) more stores; however, coupon users seem less interested in making additional store visits by 10.8%. The reshaped store-visiting tendency by gamification is maintained. Groups *Badge* and *Leaderboard* still stop by 13 and 12 more stores in the posttreatment period.

4.2. Cross-Demographic Heterogeneity

The previous results demonstrate the salient average treatments of gamification features and coupons. To understand how heterogeneous consumer segments respond to the three treatments, we conduct a series of HTE analyses across demographics, including gender, age, and income. Prior research discusses the interaction between IT adoptions and user heterogeneities (Morris et al. 2005, Venkatesh et al. 2012), such as age (Wattal et al. 2011, Huang et al. 2020), gender (Huang et al. 2018), and income (Dailey et al. 2010). As for the adoption of IT-enabled gamification, we first explore the moderating effect of gender as Koivisto and Hamari (2014) suggest that males are more likely to engage in a gamification environment. Second, we focus on the age heterogeneity given that younger generations tend to proactively adopt and participate in the IT-enabled system (Wattal et al. 2011). Finally, we consider the income effect. When shaping shopping behaviors directly, incomes may also affect shopping outcomes via gamification indirectly.

We start considering gender as the moderator of interest and extend Equation (1) by incorporating the additional interactions with $Gender_i$. The regression model is specified as

$$\begin{aligned} outcome^k_{ijt} = {} & \beta_0 + \beta_1 Badge_{ij} + \beta_2 Leaderboard_{ij} + \beta_3 Coupon_{ij} \\ & + \beta_4 Gender_i + \beta_5 Post_t + \beta_6 Badge_{ij} * Gender_i \\ & + \beta_7 Leaderboard_{ij} * Gender_i + \beta_8 Coupon_{ij} \\ & * Gender_i + \beta_9 Badge_{ij} * Post_t + \beta_{10} Leaderboard_{ij} \\ & * Post_t + \beta_{11} Coupon_{ij} * Post_t + \beta_{12} Gender_i \\ & * Post_t + \beta_{13} Badge_{ij} * Gender_i * Post_t \\ & + \beta_{14} Leaderboard_{ij} * demo^m_i * Post_t \\ & + \beta_{15} Coupon_{ij} * demo^m_i * Post_t + \alpha_1 age_i \\ & + \alpha_2 income_i + \alpha_3 iOS_i + \varepsilon_{ijt}. \end{aligned} \tag{2}$$

The main effects (e.g., $Badge_{ij}$) refer to women's responses to the treatments in the treatment period. The three two-way interactions of gender (e.g., $Badge_{ij} * Gender_i$) capture how men (i.e., $Gender_i = 1$) are different from women in reacting to gamification and couponing. The three two-way interactions of $Post_t$ (e.g., $Badge_{ij} * Post_t$) measure the posttreatment effects of women. Finally, the three-way interaction terms (e.g., $Badge_{ij} * Gender_i * Post_t$) indicate the heterogeneous posttreatment effects of men compared with women. The rest of the demographics serve as controls.

Table 7 reports the results of the HTEs regarding gender. Overall, the results are qualitatively similar across the three outcomes of interest. Women interact with both gamification and couponing, yet the extent of engagements varies much. In the treatment period, badges and leaderboards push women to spend 19% and 18.4% more, whereas coupons boost spending by 35.3%. The two-way interactions between the treatments and $Gender_i$ capture males' different behaviors compared with females. The coefficients of games and coupons are significant but have opposite directions,

Table 7. Heterogeneous Treatment Effects: Gender

	money	*distance*	*store visits*
Intercept	4.521	0.497	3.535
	(0.034)***	(0.021)***	(0.022)***
Badge	0.174	0.170	0.302
	(0.002)***	(0.002)***	(0.002)***
Leaderboard	0.169	0.180	0.291
	(0.002)***	(0.002)***	(0.002)***
Coupon	0.302	0.158	0.114
	(0.002)***	(0.002)***	(0.002)***
Gender	−0.256	−0.149	−0.235
	(0.061)***	(0.024)***	(0.021)***
Post	0.703	1.096	1.028
	(0.995)	(0.992)	(0.997)
Badge ∗ Gender	0.046	0.071	0.015
	(0.002)***	(0.002)***	(0.002)***
Leaderboard ∗ Gender	0.074	0.082	0.019
	(0.002)***	(0.002)***	(0.002)***
Coupon ∗ Gender	−0.022	−0.036	−0.021
	(0.002)***	(0.002)***	(0.002)***
Badge ∗ Post	−0.157	−0.153	−0.123
	(0.002)***	(0.003)***	(0.003)***
Leaderboard ∗ Post	−0.156	−0.165	−0.111
	(0.002)***	(0.003)***	(0.003)***
Coupon ∗ Post	−0.304	−0.145	−0.100
	(0.002)***	(0.003)***	(0.003)***
Gender ∗ Post	−0.028	−0.016	−0.020
	(0.061)	(0.024)	(0.021)
Badge ∗ Gender ∗ Post	0.031	−0.054	0.004
	(0.003)***	(0.003)***	(0.003)
Leaderboard ∗ Gender ∗ Post	0.064	−0.025	0.005
	(0.003)***	(0.003)***	(0.003)*
Coupon ∗ Gender ∗ Post	0.002	0.003	0.003
	(0.003)	(0.003)	(0.003)
Controls	Included	Included	Included
Adjusted R^2	0.650	0.522	0.449
N		16,000	

*** and * indicate 0.01 and 0.1 levels of significance, respectively.

indicating that males are motivated more by badges (24.6%) and leaderboards (27.5%) but less by coupons (32.3%) than females. In the posttreatment period, all the two-way interactions between the treatments and $Post_t$ are negative and significant. The difference between the main effect and the interactions (e.g., $Badge_{ij}$ and $Badge_{ij} * Post_t$) suggests that women become less engaged. For instance, in group *Badge*, women's enthusiasm level drops by 14.5% (i.e., $exp(-0.157) - 1$) for spending, 14.2% for strolling, and 11.6% for visiting stores. Males demonstrate the different patterns captured by the three-way interactions. After games stop, men who play badges and leaderboards continue spending 3.1% (i.e., $exp(0.031)$) and 6.6% more than women. In other words, the games reshape men's behaviors and bring additional revenues by 9.9% (compared with the males in the control group) given the persistent effects in the posttreatment period. However, they behave like the control group after the removals of monetary incentives.

Second, the gamification treatment effects could vary across different age groups as Venkatesh et al. (2012) argue that users' age moderates IT adoption. In this regard, we replace $Gender_i$ with age_i as the demographic of interest in Equation (2) and summarize the results in Table 8. The models reproduce the main findings and provide insights into the moderating effects of age. Overall, young generations are more willing to interact with games and maintain the reshaped behaviors in the posttreatment period. In contrast, couponing seems a one-for-all-age incentive. Quantitatively, in the treatment period, a 10% decrease in age promotes the badging and leaderboarding treatment effects on the amount of money spent by 1.3% (i.e., $0.9^{-0.123} - 1$) and 1.7%, on the shopping distance by 1.4% and 1.3%, and the number of store visits by 2.5% and 2.0%, respectively. These marginal effects are statistically but not economically significant in the posttreatment period compared with the baseline.

Finally, though income generally plays a substantial role in shopping behaviors (Burke 2002), how would income moderate the gamification effect? Executing the same procedure, we focus on the HTEs of income and report Table 9. We have similar patterns for the main effects, suggesting that the badge and leaderboard features engage consumers with various income levels. It is interesting to observe that a 10% income increase in group *Badge* (*Leaderboard*) leads to a 1.9% (1.9%) increase (decrease) in spending, shopping distance, and number of store visits. Thus, income amplifies the two gamification treatments in different directions. Specifically, high-income visitors are more recruited by badges, whereas the low-income population is more interested in the leaderboard feature.

Table 8. Heterogeneous Treatment Effects: Age

	money	*distance*	*store visits*
Intercept	4.493 (0.037)***	0.499 (0.021)***	3.545 (0.022)***
Badge	0.189 (0.002)***	0.172 (0.002)***	0.289 (0.002)***
Leaderboard	0.198 (0.002)***	0.180 (0.002)***	0.269 (0.002)***
Coupon	0.281 (0.002)***	0.151 (0.002)***	0.118 (0.002)***
age	0.531 (0.117)***	−0.598 (0.110)***	−0.215 (0.022)***
Post	1.037 (0.995)	1.036 (0.991)	1.039 (0.966)
*Badge * age*	−0.123 (0.005)***	−0.135 (0.005)***	−0.234 (0.005)***
*Leaderboard * age*	−0.159 (0.005)***	−0.118 (0.005)***	−0.185 (0.005)***
*Coupon * age*	−0.006 (0.005)	0.007 (0.005)	0.007 (0.005)
*Badge * Post*	−0.081 (0.003)***	−0.092 (0.003)***	−0.157 (0.003)***
*Leaderboard * Post*	−0.092 (0.003)***	−0.100 (0.003)***	−0.130 (0.003)***
*Coupon * Post*	−0.277 (0.003)***	−0.145 (0.003)***	−0.118 (0.003)***
*age * Post*	0.136 (0.117)	−0.086 (0.110)	−0.088 (0.112)
*Badge * age * Post*	−0.012 (0.006)*	−0.014 (0.006)***	−0.020 (0.006)***
*Leaderboard * age * Post*	−0.008 (0.006)	−0.008 (0.006)	−0.011 (0.006)*
*Coupon * age * Post*	0.002 (0.006)	0.003 (0.006)	0.003 (0.006)
Controls	Included	Included	Included
Adjusted R^2	0.658	0.524	0.449
N		16,000	

***, **, and * indicate 0.01, 0.05, and 0.1 levels of significance, respectively.

Table 9. Heterogeneous Treatment Effects: Income

	money	*distance*	*store_visit*
Intercept	4.473 (0.041)***	0.450 (0.021)***	3.532 (0.023)***
Badge	0.175 (0.002)***	0.227 (0.002)***	0.313 (0.002)***
Leaderboard	0.190 (0.002)***	0.243 (0.002)***	0.281 (0.002)***
Coupon	0.262 (0.002)***	0.151 (0.002)***	0.109 (0.002)***
Income	0.539 (0.115)***	0.136 (0.115)	0.254 (0.115)**
Post	0.854 (0.994)	0.925 (0.994)	0.933 (0.994)
*Badge * income*	0.200 (0.121)*	0.195 (0.121)*	0.216 (0.121)*
*Leaderboard * income*	−0.214 (0.121)**	−0.219 (0.121)*	−0.227 (0.121)*
*Coupon * income*	−0.069 (0.119)	−0.087 (0.119)	−0.095 (0.119)
*Badge * Post*	−0.092 (0.003)***	−0.108 (0.003)***	−0.196 (0.003)***
*Leaderboard * Post*	−0.097 (0.003)***	−0.166 (0.003)***	−0.115 (0.003)***
*Coupon * Post*	−0.265 (0.003)***	−0.156 (0.003)***	−0.099 (0.003)***
*income * Post*	−0.081 (0.061)	−0.032 (0.024)	−0.031 (0.021)
*Badge * income * Post*	0.183 (0.132)	−0.136 (0.132)	0.160 (0.132)
*Leaderboard * income * Post*	−0.095 (0.132)	−0.085 (0.132)	0.084 (0.132)
*Coupon * income * Post*	0.039 (0.132)	−0.079 (0.132)	0.048 (0.132)
Controls	Included	Included	Included
Adjusted R^2	0.646	0.514	0.431
N		16,000	

***, **, and * indicate 0.01, 0.05, and 0.1 levels of significance, respectively.

4.3. Within-Group Heterogeneity

We have learned that various demographics moderate the treatment effects across the experimental groups. It remains unclear to what extent the treatment effects vary within each treatment group. To shed light on the within-group heterogeneity, we conduct Levene's test and estimate a hierarchical model.

We start to analyze the equality of variances for the groups in the treatment period. Levene's test, equivalent to a one-way analysis of variance (ANOVA), calculates the within-group variances and performs joint *F*-tests. In addition, we conduct pairwise *F*-tests that detail the difference between a specific pair of groups. The results are summarized in Table 10. The joint *F*-statistics are significant, indicating that the within-group individual heterogeneity is saliently different across the treatments. Group *Leaderboard* has the highest variance followed by group *Badge*, whereas group *Coupon* has the lowest one. The results suggest that group *Leaderboard* demonstrates more polarized behaviors. As Dijkstra et al. (2008) discuss, social comparisons (i.e., competition) can be a double-edged sword that encourages or discourages people from pursuing desired behaviors. Consumers with an aggressive (laid-back) mentality may self-enforce to be more competitive (unambitious). Second, couponing is a compelling incentive for most people, explaining the mildest heterogeneity of group *Coupon*. Finally, badging is a balanced stimulus. Similar to philately, collecting badges provides the utility gain from accomplishments without the pressure

Table 10. Variance Differences

	Badge versus *Leaderboard*	*Badge* versus *Coupon*	*Leaderboard* versus *Coupon*
F-statistics	0.756***	1.072***	0.874***

Note. The outcome of interest is the amount spent.
*** indicates 0.01 level of significance.

from competition. Such a unique characteristic makes badging the most widely applied gamification feature (Raftopoulos 2015).

To sharpen the analysis of within-group heterogeneity, we adopt a hierarchical regression. One of the advantages of a hierarchical Bayes model is accounting for heterogeneity at the individual level. We consider both the group and individual levels when reconstructing the data at the individual–visit level. Focusing on the treatment effects, we specify a hierarchical model in the random coefficient manner that allows each individual to have the individual's own coefficients as

$$\begin{aligned} outcome^k_{ijt} = \beta_{0i} + \beta_{1i} Badge_{ij} + \beta_{2i} Leaderboard_{ij} \\ + \beta_{3i} Coupon_{ij} + \alpha_1 Gender_i + \alpha_2 age_i \\ + \alpha_3 income_i + \alpha_4 iOS_i + \varepsilon_{ijt}, \end{aligned} \tag{3}$$

where superscript k denotes the three outcomes by visitor i in group j at visit t. β_{0i} is the individual-specific intercept that allows for variation in the baselines. β_{1i}, β_{2i}, and β_{3i} are individual-specific slope coefficients, which jointly model the heterogeneous treatment effects on the outcomes. The demographics are included as fixed effect controls, and ε_{ijt} is idiosyncratic errors following a standard normal distribution. Because the goal is to obtain the random slopes (i.e., HTEs) at the individual level across the groups, we utilize a hierarchical Bayes approach to estimate the proposed model. The parameters in the model belong to two groups: (1) *random effect* parameters, $\boldsymbol{\beta}_i$, that vary across individuals and (2) *fixed* parameters, $\boldsymbol{\alpha}$, that do not. Thus, we can have

$$\boldsymbol{\beta}_i \sim \mathrm{MVN}(\overline{\boldsymbol{\beta}}, \boldsymbol{\Sigma}), \tag{4}$$

where $\overline{\boldsymbol{\beta}}$ denotes the mean effects that stand across individuals and $\boldsymbol{\Sigma}$ denotes the covariance matrix of $\boldsymbol{\beta}$. We benefit from conjugacy and apply Gibbs sampling to estimate the model.

We discuss the results in Table 11. Panel A reports a set of mean effects at the group level. These coefficients confirm the ATEs in Section 4.1, serving as an alternative model check. More importantly, we summarize the covariance matrix in panel B. The significance of each element on the diagonal is interpreted as the existence of the HTE for each treatment. The magnitude of a coefficient represents the degree of heterogeneity. The greater the magnitude, the higher the heterogeneity. We find that consumers are more heterogeneous in their responses to leaderboards than badges and coupons. In sum, when utilizing different modeling perspectives, ANOVA and the hierarchical model jointly characterize the distinct nature of the three treatments we study.

Table 11. Heterogeneity at the Individual Level

Panel A. Mean effects – $\overline{\boldsymbol{\beta}}$			
	money	*distance*	*store visits*
Intercept	4.390	0.508	3.537
	(0.039)***	(0.014)***	(0.023)***
Badge	0.201	0.221	0.312
	(0.002)***	(0.002)***	(0.002)***
Leaderboard	0.215	0.256	0.285
	(0.002)***	(0.002)***	(0.002)***
Coupon	0.272	0.151	0.093
	(0.002)***	(0.002)***	(0.002)***
Gender	−0.148	−0.165	−0.213
	(0.020)***	(0.019)***	(0.020)***
income	0.729	0.326	0.153
	(0.099)***	(0.114)***	(0.103)
age	0.456	−0.600	−0.738
	(0.112)***	(0.114)***	(0.121)***
iOS	0.019	0.021	0.023
	(0.012)	(0.016)	(0.016)
N		8,000	

Panel B. Within-group heterogeneity comparisons – $\boldsymbol{\Sigma}$				
	Intercept	*Badge*	*Leaderboard*	*Coupon*
Intercept	0.380			
	(0.042)***			
Badge	0.004	0.226		
	(0.057)	(0.087)***		
Leaderboard	0.009	−0.020	0.358	
	(0.044)	(0.038)	(0.014)***	
Coupon	−0.003	−0.026	−0.010	0.117
	(0.095)	(0.067)	(0.016)	(0.033)***

Notes. In panel A, the empirical standard deviations are derived from the posterior distributions of the parameters. In panel B, the outcome of interest is the shopping amount. The empirical standard deviations are derived from the posterior distributions of the parameters.

*** 0.01 level of significance.

5. Mechanism and Robustness

It is critical to identify the mechanisms behind empirics. Though prior studies explain the gamification effects using motivation theories, we empirically examine *exploration* for the observed treatment effects. Besides, we conduct two subsample analyses as robustness checks to obtain empirical regularities.

5.1. Exploration

In our off-line shopping context, the badges and leaderboards are designed to encourage visitors to purchase and stroll in the mall. The gamification groups are motivated to explore the mall. The subjects may start browsing the stores to which they have not paid attention or even make purchases. We identify each subject's newly discovered stores in the treatment period by comparing the stores visited before and after the treatment.[20] To examine the exploration effect, we reconstruct each outcome variable by calculating the ratio of the amount associated with the newly discovered stores to the total as

$$RatioNewStore_{ijt}^{k} = \frac{\text{the amount of outcome}_{ijt}^{k}\text{ at new stores}}{\text{the total amount of outcome}_{ijt}^{k}}, \tag{5}$$

where superscript k denotes the three shopping behaviors for visitor i in group j at visit t. We substitute the dependent variables in Equation (1) with *RatioNewStore* and reestimate the model.

Table 12 reports the results. The treatment effects are positive and significant, showing that the treated subjects spend and walk more at new stores than the control group. However, the coefficients of the posttreatment interactions are negative and significant, suggesting that the exploration effect melts down to some extent with the removals of the treatments. Quantitatively, the subjects who receive the badge and leaderboard treatments spend 10% and 7.6% more at the newly discovered stores in the treatment period and 0.08% and 0.07% in the posttreatment period. We also find a similar pattern regarding the other two outcomes. The results jointly suggest that gamification evokes temporal explorations because of the call-for-action nature. Considering the exploration effect here and the ATEs/HTEs earlier together, we infer that gamification not only temporarily stimulates explorations, but profoundly nourishes engagements (Vansteenkiste et al. 2006).

5.2. Subsample

One may suspect that the treatment effects are driven by contextual factors. Specifically, shopping with companions can lead to higher shopping spending, longer walking distance, and more store visits. To trace companion shopping cases, we utilize Wi-Fi tracking to detect whether a subject (i.e., a focal device) has a companion (i.e., another device) nearby during a mall visit.[21] We then conduct the subsample analysis at the visit level that excludes companion shopping occasions. Table 13 summarizes the results, wherein the robust treatment and posttreatment effects remain salient for noncompanion shopping.

Also, seasonality can bias the effects of interest. Given the experimentation in summer, the mall recruits more young visitors who are more willing to interact with games. In this regard, we reestimate Equation (1) using the subsample above the age of 22. In Table 14, gamification shows its stable, positive impacts on older generations, generalizing the findings by ruling out the concern of seasonality.

6. Conclusion and Future Research

We investigate the impact of gamification on shopping engagements. Specifically, we study badges and

Table 12. Exploration Effects

	money	*distance*	*stores*
Intercept	1.341	0.314	0.435
	(0.043)***	(0.043)***	(0.049)***
Badge	0.095	0.163	0.211
	(0.002)***	(0.002)***	(0.002)***
Leaderboard	0.073	0.296	0.304
	(0.002)***	(0.002)***	(0.002)***
Coupon	0.151	0.075	0.117
	(0.002)***	(0.002)***	(0.002)***
Post	−0.538	0.014	−1.000
	(0.981)	(0.991)	(0.995)
*Badge * Post*	−0.087	−0.101	−0.206
	(0.003)***	(0.003)***	(0.003)***
*Leaderboard * Post*	−0.066	−0.192	−0.297
	(0.003)***	(0.003)***	(0.003)***
*Coupon * Post*	−0.152	−0.077	−0.116
	(0.003)***	(0.003)***	(0.003)***
Gender	−0.238	−0.383	−0.053
	(0.058)***	(0.054)***	(0.016)***
age	0.110	0.059	−0.134
	(0.049)**	(0.057)	(0.093)
income	0.069	0.202	−0.104
	(0.074)	(0.315)	(0.057)*
iOS	0.016	0.027	−0.036
	(0.021)	(0.054)	(0.029)
Adjusted R^2	0.415	0.381	0.513
N		16,000	

***, **, and * indicate 0.01, 0.05, and 0.1 levels of significance, respectively.

Table 13. Subsamples: Noncompanion Shopping

	money	*distance*	*stores*
Intercept	4.760	0.424	3.512
	(0.044)***	(0.012)***	(0.025)***
Badge	0.174	0.207	0.307
	(0.002)***	(0.002)***	(0.002)***
Leaderboard	0.200	0.233	0.275
	(0.002)***	(0.002)***	(0.002)***
Coupon	0.269	0.137	0.095
	(0.002)***	(0.002)***	(0.002)***
Post	0.742	1.032	1.027
	(0.998)	(0.995)	(0.999)
*Badge * Post*	−0.102	−0.072	−0.077
	(0.006)***	(0.003)***	(0.003)***
*Leaderboard * Post*	−0.148	−0.106	−0.104
	(0.006)***	(0.003)***	(0.002)***
*Coupon * Post*	−0.266	−0.138	−0.095
	(0.006)***	(0.003)***	(0.003)***
Gender	−0.280	−0.130	−0.235
	(0.022)***	(0.016)***	(0.023)***
age	0.609	−0.634	−0.822
	(0.115)***	(0.114)***	(0.121)***
income	0.774	0.246	0.146
	(0.127)***	(0.117)***	(0.126)
iOS	0.038	0.028	0.021
	(0.025)	(0.015)*	(0.014)
Adjusted R^2	0.612	0.428	0.405
N		20,163	

*** and * indicate 0.01 and 0.1 levels of significance, respectively.

Table 14. Subsamples: Nonstudent Consumers

	money	*distance*	*stores*
Intercept	4.818	0.530	3.617
	(0.038)***	(0.021)***	(0.023)***
Badge	0.120	0.229	0.316
	(0.002)***	(0.002)***	(0.002)***
Leaderboard	0.199	0.243	0.308
	(0.002)***	(0.002)***	(0.002)***
Coupon	0.284	0.165	0.101
	(0.002)***	(0.002)***	(0.002)***
Post	1.044	1.040	1.006
	(0.996)	(0.996)	(0.996)
Badge ∗ Post	−0.105	−0.048	−0.086
	(0.003)***	(0.003)***	(0.003)***
Leaderboard ∗ Post	−0.153	−0.138	−0.184
	(0.003)***	(0.003)***	(0.003)***
Coupon ∗ Post	−0.256	−0.151	−0.103
	(0.003)***	(0.003)***	(0.003)***
Gender	−0.265	−0.172	−0.245
	(0.021)***	(0.021)***	(0.021)***
age	0.416	−0.581	−0.326
	(0.102)***	(0.094)***	(0.098)***
income	0.687	−0.262	0.137
	(0.115)***	(0.113)***	(0.122)
iOS	0.034	0.033	0.033
	(0.021)	(0.020)	(0.021)
Adjusted R^2	0.613	0.422	0.417
N		13,614	

*** 0.01 levels of significance, respectively.

leaderboards. Contrasted with coupons based on monetary incentives, gamification naturally pulls consumers' attention through nonmonetary games. Quantifying the effects of badges, leaderboards, and coupons on shopping behaviors, we characterize the different natures of gamification and coupons in an off-line business context. To obtain causal inferences, we conduct randomized field experiments that lead to several insights.

First, we document that gamification has a significant impact on consumer shopping engagements. Quantitatively, the average treatment effects of badges and leaderboards promote sales by 21.5% and 22.5% in the treatment period, respectively, whereas couponing delivers a more substantial promotion effect of 31.7%. Group *Badge* and *Leaderboard* subjects walk 25% and 28.4% farther than the control group for shopping distance, but the subjects in group *Coupon* walk 15.6% more only in the treatment period. Similarly, the consumers receiving the badge and leaderboard treatments tend to visit 37.9% and 33.4% more stores in the treatment period.

Second, the two-period experiment design investigates the vitality of the gamification effects. We find that game features nourish long-term shopping behaviors even after the games stop, whereas the effect of coupons fades out dramatically once coupons expire. We specifically show that in the posttreatment period, the consumers in groups *Badge* and *Leaderboard* continue spending $11.26 and $8.26 more compared with the control group. Gamification also continues the engaged behaviors regarding the other two outcomes of interest.

Third, we study how heterogeneous consumers respond to gamification and coupons. We start with the moderating effects of demographics and then adopt a hierarchical Bayes model to sharpen the analysis of within-group heterogeneity. The results suggest that coupons are more attractive to women, whereas gamification is more interesting to men. Males remain energetic in games than their counter-gender in the long run. Young generations seem more engaged with the games. Our results also indicate that gamification is a general tool to motivate both high- and low-income consumers. Besides, group *Leaderboard* reveals the highest heterogeneity in the within-group heterogeneity analysis. Social comparisons on leaderboards can be a double-edged sword, wherein more (less) engaging subjects develop a self-reinforcing (self-banishing) process. Unlike competing on leaderboards, collecting badges with a moderate within-group heterogeneity provides a more balanced strategy to engage the general public.

Finally, we conduct mechanism checks to examine the exploration effects driven by gamification. After identifying consumers' newly discovered stores in the treatment period, we find that consumers in groups *Badge* and *Leaderboard* spend and walk more at new stores than the control group. However, such an exploration effect fades out after the removals of the treatments. Thus, gamification evokes a temporal exploration effect because of their call-for-action nature. In addition, to address the companion and seasonality effects, our subsample analyses find that gamification shows its stable impacts on noncompanion shopping and nonstudents shoppers.

This study contributes to the literature in the following ways. First, we complement prior studies by exploring a new business context. We construct a novel online-to-off-line gamified environment to study consumer engagements in a brick-and-mortar mall. We also extend the scope of the gamification literature by simultaneously comparing multiple game elements in a single setting and benchmarking them against monetary incentives. Second, we conduct randomized field experiments to address the call for causal inferences (Hamari et al. 2014). Thousands of general shoppers are randomly assigned to stimuli in a well-controlled experiment. Third, to the best of our knowledge, we are the first to examine the gamification effect with a broader time horizon. Specifically, our study considers both treatment and posttreatment periods to capture potential behavior changes after gamification stops. Fourth, we quantify the heterogeneous gamification effects while considering the moderations of demographics generally ignored in the literature. We further compare the within-group heterogeneities across gamification and couponing. Finally, this study showcases the potential of location-based gamification to engage consumers. When extant literature focuses on nonlocation-based applications (e.g., Denny et al. 2018, Shao et al. 2019), we show how gamification evolves with the new development of technologies. This study helps better understand the effectiveness of a new variant of gamification, which tracks and incentivizes users via location-based technologies.

This study brings useful managerial implications to mall managers and gamification designers. For mall management, the findings help tailor effective gamification strategies to promote desirable shopping engagements. Although couponing is a straightforward method to boost sales immediately, gamification not only grows revenues in a short-term manner but, more importantly, reshapes consumer shopping patterns in a long-term fashion. Badges and leaderboards can be more effective than coupons in calling for actions, such as explorations. Though not directly contributing to sales, these nonpurchase engagements can serve as a mediating step to generate additional interactions. The reason behind this may be that consumers are encouraged to explore the mall, discover newly opened stores, and try new product arrivals. In this regard, the pull model of gamification promotes the matches between consumers' interests and stores' selections. Besides, the HTE results help develop engagement strategies using game features. Gamification is relatively preferred by males, whereas coupons are generally more attractive to females. A leaderboard can be implemented to engage men given their competing nature. Following a similar logic, we could leverage games to draw the attention of youths. These gamification-driven engagements should enhance shopping experiences and satisfaction. Finally, the management should consider consumer heterogeneity to perform mass marketing or segmentation. Badging is a perfect choice to call for purchases from the entire pool. In contrast, a leaderboard is particularly useful to extract the surplus from the most self-motivated players. We expect more compelling use cases as long as the desired outcomes are well-incorporated into games. Overall, such location-based gamification provides several tools for mall managers to run more interactive promotion strategies.

For gamification designers, it is critical to align incentive designs and outcomes of interest (Liu et al. 2017). To have suitable matches, designers should fully unlock the potentials of gamification that can be programmed flexibly. Take badging as an example. A designer can entice consumers to visit specific stores by crafting more eyeball-catching badges and putting these badges on the salient positions in the program. Similarly, offering badge superbuyers should lead to more purchases. Besides, a leaderboard's point system should be delicately configured for a fair competing environment. Without a fair game, users get disappointed quickly and lose their passion for engaging. Finally, designers may try to deliver more dynamic game components, such as real-time treasure hunt events to enhance game experiences using the latest location-based technologies (e.g., Wi-Fi tracking).

As taking an initial step to study gamification in an off-line shopping context, this study is constrained by experimentation-specific factors. First, we acknowledge that game design plays a significant role in driving the gamification effects. Ideally, game features call for the same action, yet we study badges and leaderboards by rewarding consumers based on different goals. Though the two are imperfectly paired, the correlation alleviates such worries to some extent. Meanwhile, we acknowledge this issue as a limitation and a priority in future directions. Second, we currently consider badging and leaderboarding, but there exist other innovative game elements, such as progress bars. A natural way to extend this research is to examine more features, the associated effects, and the mechanisms behind them. Third, our coupons are store-specific because of the limited experiment resources. The ideal coupon design should be mall-wide to avoid store-specific preferences. We hope to implement such ideal coupons in an unlimited experiment setting in a future study. Finally, we contrast gamification with couponing but have not studied the complementarity and substitution between the two. We would contribute to the literature by filling this research gap. Despite the aforementioned limitations and new directions, this study, to the best of our knowledge, is among the first to document the robust gamification effects on off-line shopping engagements. Overall, this study generates executable guidance on how businesses effectively leverage such a novel IT artifact to harvest desired outcomes.

Acknowledgments

The authors are listed in alphabetical order with equal contribution. The authors thank the senior editor, the associate editor, and three anonymous reviewers for constructive suggestions. The authors also thank the participants at the 2018 Workshop on Information Systems and Economics, the 2019 Symposium on Statistical Challenges in Electronic Commerce and Research, and the 2019 Conference of Information Systems and Technologies as well as research seminars at the University of Texas at Dallas, Rutgers University, the University of Tennessee at Chattanooga, National Taiwan University, and National Chung Cheng University for helpful comments.

Appendix. Survey Questionnaire

We conduct a brief survey after the random assignments. The purpose of this postassignment is to collect consumers' demographics that can be used to perform the postassignment balance check. It takes about one minute to finish the survey. The system informs consumers that they will automatically enter a drawing for a $50 gift card lottery if fulfilling the survey completely. To control the quality of the responses, we also declare that the prizes can be collected only if the inputted data match the winners' identifications. The winners were notified on September 1, 2018 (i.e., one week after the end of the experiment). We attach the questionnaire for group *Badge* as follows.

(1) Please rate your satisfaction with the overall shopping experience recently? (disappointed–satisfied)

(2) Please confirm whether you are successfully enrolled in the special program. (yes/no)

(3) Please provide your information using the scroll bars.

- Your gender
- Your age (16–60)
- Your annual income level (500–60,000)

(4) Please indicate whether you are shopping alone. (yes/no)

Endnotes

[1] More than 90% of mall visitors take advantage of the free Wi-Fi access.

[2] Badges have no monetary values and are only visible to consumers themselves.

[3] "Badge" and "badging," "leaderboard" and "leaderboarding," "coupon" and "couponing" are interchangeable throughout.

[4] Several applications integrate game-like elements into nongame contexts, such as the point systems for frequent flyer programs and membership status. Gamification yet differentiates itself from these loyalty programs using additional motivations (e.g., social recognitions) beyond expenditure-based metrics (Huotari and Hamari 2012).

[5] The shopping mall remains unnamed in this study per the nondisclosure agreement.

[6] We approximate the number of visitors utilizing the Wi-Fi service using the number of unique IP addresses logging onto the Wi-Fi system. The number is 89.73% of the number of daily visits on average.

[7] Because of the computation limitation, we randomly select 20% of the effective subjects for the following analyses.

[8] This autoenrollment procedure is designed to prevent the subject from self-selecting to receive the treatment. The design is implemented to quantify the average treatment effects, wherein all subjects are enrolled to the corresponding groups.

[9] There are nine store categories in total, including books, cosmetics, drinks, dining, electronics, fashion, groceries, movies, and toys.

[10] There is no monetary value associated with badges.

[11] A visitor earns one point for every 0.06 mile and 20 points for every store visit.

[12] The accumulated points are only used for promoting participants' ranks and cannot be exchanged for money or discounts.

[13] A lottery ticket is given to incentivize a visitor to fulfill the survey. The questionnaire of the survey is attached in the appendix.

[14] This task can be achieved when mobile payments (e.g., WeChat Pay and Alipay) are used. For the visitors not using mobile payments, sales representatives simply look up membership accounts.

[15] We define a store visit as staying a store more than three minutes.

[16] This information is directly collected from the Wi-Fi system.

[17] Lowercases are used to indicate the logarithm values of the controls.

[18] This also provides a fair comparison between the two periods by dispelling the doubt of unexpected shocks.

[19] Group *Control* strolls 1.85 miles and visits 34.81 stores on average.

[20] We define the stores visited two months before the experiment as the already known stores. Then, we identify each subject's fixed set of the newly discovered stores in the entire experiment window (i.e., eight weeks) by excluding the subject's already known stores.

[21] We also use the input from the survey in the appendix to verify the Wi-Fi detection procedure. The two data sources verify each other and have a 92% overlap.

References

Abramovich S, Schunn C, Higashi RM (2013) Are badges useful in education? It depends upon the type of badge and expertise of learner. *Ed. Tech. Res. Dev.* 61(2):217–232.

Amo LC, Liao R, Rao HR, Walker G (2018) Effects of leaderboards in games on consumer engagement. *Proc. 2018 ACM SIGMIS Conf. Comput. People Res.* (Association for Computing Machinery, New York), 58–59.

Antin J, Churchill E (2011) Badges in social media: A social psychological perspective. *CHI 2011 Gamification Workshop Proc.* (Association for Computing Machinery, New York), 7(S2).

Biles M, Plass JL, Homer B (2018) Designing digital badges for educational games: The impact of badge type on student motivation and learning. *Internat. J. Gaming Comput. Mediated Simulations* 10(4):1–19.

Bittner JV, Shipper J (2014) Motivational effects and age differences of gamification in product advertising. *J. Consumer Marketing* 31(5):391–400.

Bowey JT, Birk MV, Mandryk RL (2015) Manipulating leaderboards to induce player experience. *Proc. 2015 Annual Sympos. Comput. Human Interaction Play* (Association for Computing Machinery, New York), 115–120.

Burke R (2002) Technology and the customer interface: What consumers want in the physical and virtual store. *J. Acad. Marketing Sci.* 30:411–432.

Buunk AP, Gibbons FX (2007) Social comparison: The end of a theory and the emergence of a field. *Organ. Behav. Human Decision Processes* 102(1):3–21.

Buunk BP, Collins RL, Taylor SE, VanYperen NW, Dakof GA (1990) The affective consequences of social comparison: Either direction has its ups and downs. *J. Personality Soc. Psych.* 59(6): 1238–1249.

Cavusoglu H, Li Z, Huang K-W (2015) Can gamification motivate voluntary contributions? The case of StackOverflow Q&A community. *Proc. 18th ACM Conf. Companion on Computer Supported Cooperative Work & Social Computing* (Association for Computing Machinery, New York), 171–174.

Christy KR, Fox J (2014) Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance. *Comput. Ed.* 78:66–77.

Costa JP, Wehbe RR, Robb J, Nacke LE (2013) Time's up: Studying leaderboards for engaging punctual behavior. *Proc. First Internat. Conf. Gameful Design, Research, and Applications* (Association for Computing Machinery, New York), 26–33.

Dailey D, Bryne A, Powell A, Karaganis J, Chung J (2010) Broadband adoption in low-income communities (Social Science Research Council, New York).

Deci E, Ryan RM (1985) *Intrinsic Motivation and Self-Determination in Human Behavior* (Springer US, New York).

Denny P, McDonald F, Empson R, Kelly P, Petersen A (2018) Empirical support for a causal relationship between gamification and learning outcomes. *Proc. 2018 CHI Conf. Human Factors in Computing Systems* (Association for Computing Machinery, New York), 1–13.

Deterding S, Dixon D, Khaled R, Nacke L (2011) From game design elements to gamefulness. *Proc. 15th Internat. Academic MindTrek Conf.: Envisioning Future Media Environments* (Association for Computing Machinery, New York), 9–15.

Dijkstra P, Kuyper H, Van der Werf G, Buunk AP, van der Zee YG (2008) Social comparison in the classroom: A review. *Rev. Ed. Res.* 78(4):828–879.

Domínguez A, Saenz-De-Navarrete J, De-Marcos L, Fernández-Sanz L, Pagés C, Martínez-Herráiz J-J (2013) Gamifying learning experiences: Practical implications and outcomes. *Comput. Ed.* 63:380–392.

Duggan K, Shoup K (2013) *Business Gamification for Dummies* (John Wiley & Sons, Hoboken, NJ).

Festinger L (1954) A theory of social comparison processes. *Human Relations* 7(2):117–140.

Fotaris P, Mastoras T, Leinfellner R, Rosunally Y (2016) Climbing up the leaderboard: An empirical study of applying gamification techniques to a computer programming class. *Electronic J. e-Learn.* 14(2):95–110.

Hamari J (2013) Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Res. Appl.* 12(4):236–245.

Hamari J (2017) Do badges increase user activity? A field experiment on the effects of gamification. *Comput. Human Behav.* 71: 469–478.

Hamari J, Koivisto J, Sarsa H (2014) Does gamification work? A literature review of empirical studies on gamification. *47th Hawaii Internat. Conf. System Sci.* (IEEE, Piscataway, NJ), 3025–3034.

Hanus MD, Fox J (2015) Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Comput. Ed.* 80:152–161.

Hassan L (2016) Governments should play games: Towards a framework for the gamification of civic engagement platforms. *Simulation Gaming* 48(2):249–267.

Hofacker CF, de Ruyter K, Lurie NH, Manchanda P, Donaldson J (2016) Gamification and mobile marketing effectiveness. *J. Interactive Marketing* 34:25–36.

Huang B, Hew K (2015) Do points, badges and leaderboard increase learning and activity: A quasi-experiment on the effects of gamification. *Proc. 23rd Internat. Conf. Comput. Ed.* (Society for Computer in Education), 275–280.

Huang N, Burtch G, Hong Y, Pavlou PA (2020) Unemployment and worker participation in the gig economy: Evidence from an online labor market. *Inform. Systems Res.* 31(2):431–448.

Huang N, Burtch G, Gu B, Hong Y, Liang C, Wang K, Fu D, Yang B (2018) Motivating user-generated content with performance feedback: Evidence from randomized field experiments. *Management Sci.* 65(1):327–345.

Huotari K, Hamari J (2012) Defining gamification: A service marketing perspective. *Proc. 16th Internat. Academic MindTrek Conf.* (Association for Computing Machinery, New York), 17–22.

Hydari MZ, Adjerid I, Striegel A (2019) Health wearables, gamification, and healthful activity.

Koivisto J, Hamari J (2014) Demographic differences in perceived benefits from gamification. *Comput. Human Behav*. 35:179–188.

Kwon HY, Özpolat K (2020) The dark side of narrow gamification: Negative impact of assessment gamification on student perceptions and content knowledge. *INFORMS Trans. Ed.* 21(2):67–81.

Kwon KH, Halavais A, Havener S (2015) Tweeting badges: User motivations for displaying achievement in publicly networked environments. *Cyberpsychology Behav. Soc. Networking* 18(2):93–100.

List JA, Rasul I (2011) Field experiments in labor economics. Ashenfelter O, Card D, eds. *Handbook of Labor Economics* (Elsevier), 103–228.

Liu D, Santhanam R, Webster J (2017) Toward meaningful engagement: A framework for design and research of gamified information systems. *Management Inform. Systems Quart.* 41(4):1011–1034.

Mollick E, Rothbard N (2014) Mandatory fun: Consent, gamification and the impact of games at work. *Ethics eJournal*.

Morris MG, Venkatesh V, Ackerman PL (2005) Gender and age differences in employee decisions about new technology: An extension to the theory of planned behavior. *IEEE Trans. Engrg. Management* 52(1):69–84.

Patten B (2016) Effective training strategies: 7 companies using gamification correctly. Accessed December 9, 2021, https://blog.insynctraining.com/effective-training-strategies-7-companies-using-gamification-correctly.

Peng W, Lin J-H, Pfeiffer KA, Winn B (2012) Need satisfaction supportive game features as motivational determinants: An experimental study of a self-determination theory guided exergame. *Media Psych.* 15(2):175–196.

Picone B, Mundorf J, Davidsson G (2019) Gymification-inducing intrinsic motivation through a gamified gym application. Dissertation, accessed December 9, 2021, http://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-44068.

Raftopoulos M (2015) How enterprises play: Towards a taxonomy for enterprise gamification. *Proc. DiGRA* (Digital Games Research Association).

Randall DL, Harrison JB, West RE (2013) Giving credit where credit is due: Designing open badges for a technology integration course. *TechTrends* 57(6):88–95.

ReportLink (2019) Global gamification market by solution, by deployment, by organization size, by application, by end-user vertical, by region, competition, forecast & opportunities, 2024. Accessed December 9, 2021, https://www.reportlinker.com/p05762137/Global-Gamification-Market-By-Solution-By-Deployment-By-Organization-Size-By-Application-By-End-User-Vertical-By-Region-Competition-Forecast-Opportunities.html?utm_source=PRN.

Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Amer. Psych.* 55(1):68–78.

Sailer M, Hense J, Mandl H, Klevers M (2013) Psychological perspectives on motivation through gamification. *Interaction Design Architecture(s) J.* 19:18–37.

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

Shang S, Lin K (2013) An understanding of the impact of gamification on purchase intentions. *AMCIS 2013 Proc.* (Association for Information Systems), 11.

Shao Z, Zhang L, Zhang R, Pan Z (2019) Impact of gamification on consumers' online impulse purchase: The mediating effect of affect reaction and social interaction. *PACIS* (Association for Information Systems), 21.

Sigala M (2015) Gamification for crowdsourcing marketing practices: Applications and benefits in tourism. Garrigos-Simon F, Gil-Pechuán I, Estelles-Miguel S, eds. *Advances in Crowdsourcing* (Springer International Publishing, New York), 129–145.

Teixeira PJ, Silva MN, Mata J, Palmeira AL, Markland D (2012) Motivation, self-determination, and long-term weight control. *Internat. J. Behavioral Nutrition Physical Activity* 9(1):22.

Tobon S, Ruiz-Alba JL, García-Madariaga J (2020) Gamification and online consumer decisions: Is the game over? *Decision Support Systems* 128:113167.

Vansteenkiste M, Lens W, Deci EL (2006) Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Ed. Psyc.* 41(1):19–31.

Venkatesh V, Thong JYL, Xu X (2012) Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *Management Inform. Systems Quart.* 36(1):157–178.

Wang L, Gunasti K, Shankar R, Pancras J, Gopal R (2020) Impact of gamification on perceptions of word-of-mouth contributors and actions of word-of-mouth consumers. *Management Inform. Systems Quart.* 44(4):1987–2011.

Wang X, Sanders G (2019) For money, and for fun: Exploring the effects of gamification and financial incentives on motivating online review generation. *AMCIS* (Association for Information Systems).

Wattal S, Hong Y, Mandviwalla M, Jain A (2011) Technology diffusion in the society: Analyzing digital divide in the context of social class. *2011 44th Hawaii Internat. Conf. System Sci.*, 1–10.

Wehrens MJPW, Kuyper H, Dijkstra P, Buunk AP, Van Der Werf MPC (2010) The long-term effect of social comparison on academic performance. *Eur. J. Soc. Psych.* 40(7):1158–1171.

Wells BM, Skowronski JJ (2012) Evidence of choking under pressure on the PGA tour. *Basic Appl. Soc. Psych.* 34(2):175–182.

Williams GC, Niemiec CP, Patrick H, Ryan RM, Deci EL (2009) The importance of supporting autonomy and perceived competence in facilitating long-term tobacco abstinence. *Ann. Behav. Med.* 37(3):315–324.

Wu Y, Kankanhalli A, Huang K-W (2015) Gamification in fitness apps: How do leaderboards influence exercise? *ICIS*.

Xi N, Hamari J (2020) Does gamification affect brand engagement and equity? A study in online brand communities. *J. Bus. Res.* 109:449–460.

Zhang Y, Van Horen F, Zeelenberg M (2021) Increasing saving intentions through leaderboards: A gamification approach. *PLoS One* 16(4):e0249283.