Efficiency of Digital Enabled Business Platforms: a Perspective from Understanding User Behavior

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Zhuojun Gu

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisors: Ravi Bapna, Alok Gupta

April 2017

Acknowledgements

My 6 years of Ph.D. study in the Information & Decision Sciences (IDSc) Department of University of Minnesota has been a laborious but really valuable experience. I would like to take this chance to express my gratitude to all the people around that have supported me in accomplishing my Ph.D. degree. Without their help, I can hardly imagine how I would successfully fulfill my graduate trainings and this dissertation.

I am most grateful to my co-advisors, Prof. Alok Gupta and Ravi Bapna. Since the beginning of my Ph.D. program, they systematically taught me how to identify research questions and solve problems in department seminar courses and through regular meetings. I find myself benefit most from their way of forming research questions. I always bear in mind Alok's advice that although methodology is important, the ability to find and solve interesting research problems will determine one's ultimate contribution to the academic area. Ravi impressed me with his ability to keep up with the ever changing research frontier driven by the rapid technology advances. I feel lucky to work with mentors with such great enthusiasm and originality. Ravi and Alok are not only mentors on my research approach, but also great supporters that cheer me up emotionally when my research reach an impasse.

I am very thankful to Prof. Jason Chan, who joined our IDSc department as a. faculty member during my last year in University of Minnesota. We had many constructive discussions over the design of the randomized field experiment of the second essay in this dissertation. Jason also offered substantial supports in preparing me for the job market.

I would also like to acknowledge the very kind help from Prof. Yuqing Ren, Prof. Gerard McCullough and Prof. Terrance Hurley. Prof. Ren directed me to a very important body of behavioral literature that relates closely to my research topic. I took my Ph.D. level Econ classes from Prof. Gerard McCullough and Prof. Terrance Hurley. Their classes lay a sound foundation for the econometric modeling techniques I use in this dissertation. Also, I appreciate their time for being my external committee members in the oral defense and dissertation defense.

I want to especially thank Mochen Yang, a Ph.D. student in our IDSc program, and Mahdieh Ta, a research assistant in IDSc department, for their technical support in data retrieval and collection.

My cohorts in the Carlson School of Management provide a very mutual supportive environment. Thanks to Dr. Jingjing Zhang who shared with me her research and job market experience as a student senior to me. Thanks to Ali Mahdavi, my very considerate office mate, who provided me with great convenience and comfort during my family emergency in year 2012 and 2013. Thanks to Zuhui Xiao, a Ph.D. student in the marketing department and my best friend in Carlson, provide lots of beneficial feedback through our daily conversation on research topics. Thanks to admin Kat Nelson and former Ph.D. director Earlene Bronson who make my life easy during the Ph.D. program.

Finally, this dissertation is dedicated to my grandparents and parents who brought me up and cultivated my desire for knowledge since my early age. I always feel that I can never sufficiently express my gratitude toward them for their unconditional support along my way to achieve the PhD degree.

This dissertation is composed of two essays in co-authorship with Prof. Ravi Bapna, Prof. Alok Gupta and Prof. Jason Chan. All data sources for the essays are obtained either through my advisors' industry connection or through experiments supported by my co-authors' research funding. And Zhuojun Gu is solely responsible for the analysis and contents of this publication.

.

Table of Contents

List of Tables

List of Figures

Overview

Advancements in information technology give birth to a wide variety of digital enabled business platforms. Commonly used online platforms range from online auction sites to mobile gaming apps. Users' bidding strategies and gaming behavior under these particular online settings are quite different from those under the traditional scenarios. This dissertation empirically examines user behavior in two digital enabled platform scenarios, namely, the eBay auctions site and a social mobile game app. I aim to understand the impact of online user behavior on the platform efficiency as well as how platform settings affect usage outcome.

Broadly, there are two ways that online settings may affect platform efficiency through changing users' behavior. The first way is by directly influencing the transaction rules or users' usage experience on the platform. And users respond by adapting their strategy or usage pattern. In another word, the impact of online setting on platform efficiency is channeled through the bi-directional interaction between the user and the platform. For example, Ghose et al (2014) discuss the impact of in-app purchase and in-app advertising settings on user demand. The second way for the online setting to affect platform efficiency is by influencing users' interaction with her peer users of the platform. This is a richer scenario where the overall effect of the online setting over the user community is taken into consideration. In Oestreicher-Singer et al (2013), the authors discussed how the social computing features affect users' willingness to pay for content providing platforms.

I include both the first and the second mechanism when pursuing my research questions. In my first essay, I address the impact of bidders' use of the sniping strategy on the efficiency of the eBay auction site by looking at the bidder welfare. Sniping is defined as a bidding strategy that bidders hold their bids till the last 15 seconds before the end of an auction. I consider the efficiency of the auction platform to be determined not only by the interaction between the bidder and the auction platform, but more an equilibrium outcome of the sniper and all other competing bidders. Snipers are rational bidders who adapt to the unique proxy bidding setting of eBay while taking into account the impact of other early bidders. Since the concurrence of early proxy bidding by a mass audience may lead to price war for common value goods that increase the ending price and compromise bidder welfare, snipers postpone submitting their bids till the very end of an auction to avoid the winner's curse (Bijari and Hortacsu, 2003). Prior literature formally discussed the various motivations that drive the use of sniping strategy on a theoretical basis. For example, Bijari and Hortacsu, (2003) justify sniping as a gain-maximizing strategy employed to avoid price war for common value products. Roth and Ockenfels (2000) find that bidders submit late bids to prevent transaction cost under certain minimum increment bidding rules.

However, due to the many confounding factors that perturb the equilibrium condition, to what extent does sniping strategy benefit bidders is a question subject to empirical examination. The main challenge for the empirical quantification of the returns to sniping lies in users' endogenous use of the sniping strategy. The use of the sniping strategy may be endogenous to the bidder's value type. The endogeneity may also arise from users' selection of auctions to snipe. We measure the bidder welfare, or the returns to sniping by the bidder surplus, which is the difference between bidders' valuation and the price paid. However, the highest bid of an auction is not revealed to the public due to the proxy bidding setting on eBay. Thus, in addition to the endogeneity issue mentioned above, the data challenge is compounded by the lack of observable bidder value. We solve these data challenges by obtaining a proprietary dataset of Xbox 360 consoles on eBay that contains the full bidding history for all bidders over a period of 5 years and tackle the endogeneity issue by applying the 3SLS regression technique. Contrary to our expectation, the study showed that sniping actually significantly raise rather than lower the auction ending prices. However, the winner surplus of snipers are still higher than early bidders since snipers are high type bidders whose valuation of an auction item is higher than that of the early bidders.

In my second essay, I investigate the platform efficiency of a mobile gaming app by looking at its ability to engage and retain players. I particularly explore how the game app user engagement and retention level is influenced by adding the crowdsourcing feature to the mobile gaming apps. User engagement and retention are two important indices for users' usage outcome. While engagement emphasizes the intensity of user participation, retention focuses on the endurance and sustainability of user's usage behavior.

I expect the crowdsourcing feature that involves players in the game design process to motivate user interaction with both the game platform and the player community. On the bi-directional individual-platform interaction level, we propose that the crowdsourcing features enhance user engagement and retention by allowing users to submit content to the game platform as well as to access the crowdsourced content. First, according to the "Ikea Effect" and the task design theory, incorporating user contribution in products can increase users' sense of ownership or autonomy. For the gaming app context, I expect that effort-based investment in the form of content submission, as well as the control over the game through the option to submit content will lead to users' commitment to the game app, which foster user engagement and extends user retention. Second, access to crowdsourced content that brings more novelty to the game will help to trigger users' interest in the game and sustain their interest for a longer period of time. Thus, access to crowdsourced content, as an integral sub-option of crowdsourcing is also considered effective in leveraging user engagement and retention. On the community level, the interaction among users of the player community is strengthened by the crowdsourcing feature that connects users through the commonly interested topic that users contribute to. Users implicitly communicate and resonate with other players if they find the game content contributed by other players similar to their taste and get aware that their submitted content will be recognized by other players. Due to the stronger attachment between the individual user and the user community, the crowdsourcing feature setting increases user engagement and retention level.

Since user's participation in the crowdsourcing options is endogenous to their original engagement and retention level, we cannot estimate the causal impact of crowdsourcing features on these usage outcomes through an observational study. To establish the casual relationship between the crowdsourcing feature and the usage outcome, we thus partner with a social mobile game company to launch a clean and neat randomized field experiment. We compare the user engagement and retention level between game users treated with the crowdsourcing feature and users from the control group. Further, we try to discover through what mechanism the crowdsourcing feature affects users' usage outcome.

Main findings of the second essay include that both the content submission treatment and the access to crowdsourced content treatment significantly enhance user retention, but only content submission treatment significantly enhances user engagement. Besides, we find that the engagement and retention improvement can be caused by mere exposure to the content submission option. It shows that without actual submission behavior, mere exposure to content submission option still boost user engagement and retention by increasing users' perceived control and autonomy over the game app. On the contrary, users treated by the access to crowdsourced content option show improvement in user retention only when they are actually exposed to the crowdsourced content. In another word, the retention improvement is not based on users' awareness of the access to crowdsourced content option. Much to our surprise, user engagement gets significantly improved under the content submission treatment while the user engagement level does not show significant improvement under the treatment when both the content submission option and the content access option are provided. A possible explanation can be that users under the treatment where only content submission option is available have more perceived control than the users under full-crowdsourcing treatment. As users are only exposed their own submitted content in the submission-only group, they are not distracted by the random crowdsourced content from others and the submission-only version has higher customizability. It is this high perceived control that encourages engagement. Finally, we find users show the greatest improvement in user retention when they both submit content and are exposed to crowdsourced content. We also provide evidence that crowdsourcing features impose heterogeneous impacts on different user segments.

In both essays of my dissertation, I try to understand how user behavior under certain online settings affects the platform efficiency, including how user sniping strategy, enabled by the online proxy bidding setting, affects bidder surplus and how crowdsourcing features change the gaming app’s ability to engage and retain users. I will include more details about the two studies in the remaining sections of this dissertation.

Essay 1 – Examining Returns to Sniping on eBay

1. Introduction

Accepted conventional wisdom is that there are significant benefits to sniping, bidding in the last 15 seconds of an auction, on eBay (Roth and Ockenfels 2002). However, there are few, if any, empirical studies that have examined the benefits of sniping. Knowing the economic returns to sniping on eBay has important implications for buyer welfare and seller efficiency.

The issue of measuring returns from sniping is not straight forward. For example, suppose we could establish that sniping is associated with higher bidder surplus. However, the reason for increased surplus could be either that the snipers have larger valuations or it may be that snipers only participate in auctions that are likely to end in lower prices. Therefore, the impact of sniping on bidders cannot be simply answered by measuring surplus. It is in fact endogenous to the bidder's type as well as the auction dynamics. Indeed since earlier research has established that sniping is endogenous (Bajari and Hortacsu, 2003, Roth and Ockenfels, 2000), the real interesting question with respect to sniping's impact is linking it to the overall welfare impacts on the sellers as well as developing nuance on how it relates to our knowledge of bidder's strategies. To our knowledge, our study is the first one to examine and answer these questions.

Prior literature has provided variety of theoretical justifications for sniping in eBay due to the usage of hard closing time for auctions, unlike traditional open-cry auctions that have a "popcorn" style closing. These theoretical justifications range from avoiding bidding wars with incremental or like-minded bidders to informational gains sought by experts who don't want to give other bidders the opportunity to respond to their bids.

Since eBay forces all bidders to use its proxy-bidding agent and implements a second-price equivalent mechanism; therefore, theoretically, there is no need for strategic bidding on part of bidders. Thus, at some level, the presence of sniping can also be interpreted as a general misunderstanding of eBay's proxy bidding scheme – however, the claim has to be made with caution as eBay, because of the ability of bidders to revise their bids it is not a pure sealed-bid second price auction.

While not much is explored in terms of returns to sniping, Gray and Reiley (2007), in a controlled field experiment of 70 very similar or identical auctions (with respect to seller, item characteristics), find no significant returns to sniping. In the experiment, they randomly use the proxy bidder in some and snipe the others. However, the experimental setting assumes that sniping is an exogenous act and not endogenous as other researchers have found. The consideration for endogeneity is a key contribution of our research. A key challenge for researchers attempting to use publicly available eBay data to answer the return question is the inability to find which is essentially indexed as the highest bid placed (Bapna et al, 2008a), which is not available to the researchers. By virtue of a proprietary dataset obtained, we isolate the price impact of sniping from the price impact of being a sniper. Since bidder's value type is directly acquired from eBay's data warehouse that provides us with full bidding histories across year 2006-2010 of Xbox360 game console auctions, including the highest bid made by the winning bidder, we overcome this limitation. We econometrically examine the price paid by winners who used sniping strategy against those who did not, where sniping is instrumented by the value type of the winner and observed and unobserved bidder, seller and auction characteristics are controlled.

Our econometric model also features the interdependence between auctions. We include the number of auctions overlapping in time and the number of cross bidders as explanatory variable, which has significant implication for buyer welfare. Traditional auction theories view one auction as unit of analysis. However, this gives an incomplete picture of online bidding scenario. Online auctions don't take place in isolation. According to our data, for every online auction, on average 8 other competing auctions are held simultaneously.

Our analysis results in the following novel and interesting findings:

1) Auctions that attract snipers realize significantly higher prices than those that don't, and that snipers are higher value type bidders that also end-up realizing higher surplus.

2) While it is well known that the number of bidders is positively correlated with the ending price, we find that the auctions with lower number of bidders and those that overlap with fewer other auctions are more likely to attract snipers, thereby shedding light on endogenous nature of sniping;

3) Interestingly, we find that the well-established main effect of the impact of seller reputation on price disappears when the number of bidders is treated as being endogenous. The positive effect of seller reputation on price is fully mediated through the number of bidders attracted by the auction. In another word, higher seller reputation increases the competition intensity of the auction rather than directly increase the buyers' willingness to pay;

4) A sub-sample analysis of the high vs. low priced versions of the product under consideration indicates that strategic behavior increases as the stakes get higher.

2. Literature Review

There is a significant literature on sniping behavior as well as on seller and buyer efficiency in eBay auctions. Bijari and Hortacsu (2003) tailored Milgrom and Weber's (1982) common value goods auction model for eBay auction, in which the equilibrium bidding strategy focuses on the timing of entry rather than biding price due to the hard closing time of eBay auction. They showed that under the setting of common value goods and second price auction, all bidders have the incentive to hide their private information till the last minute. Thus, they justified the use of sniping as a dominant strategy to avoid price war for common value product. Roth and Ockenfels (2000), by comparing eBay (with hard closing time) with Amazon (sequential) auctions, and computer with antiques as auction items, found out that aside from common value, multiple causes leads to one's late submission of true value, such as expected transaction price as a result of initial bid and minimum increment.

Traditionally, one measure that captures the consumer welfare in online market is consumer surplus. As the final payment in a second price auction is independent of the winner's own valuation, the difference between the true value and the price paid constitute consumer surplus. Due to the well-known result that second-price sealed auction encourages the bidders to bid their true value (Vickrey, 1962), the highest bid of the bidder could be viewed as the true valuation. Based on this, Bapna et al (2008a) deployed a sniping agent to estimate overall consumer surplus levels on eBay. However, due to the nature of their dataset, they do not have information about auctions that were won by non-snipers, which did not allow them to analyze the price impact of sniping. Hu and Alvaro (2008) used large variety eBay product transaction dataset to show that sniping lead to high surplus ratio. By defining surplus ratio as surplus over final value, the authors notice the heteqrogeneity in consumer valuation. However, they were still unable to identify whether there is a significant differences among the bidder types (in terms of their valuations) between snipers and non-snipers. Finally, increasingly researchers are realizing that looking at auctions in isolation without considering the other simultaneous or overlapping auctions for same items does not fully capture strategic actions of bidders. For example, Bapna et al (2009) discovered that overlap of an auction with other competing auctions has a significant negative influence on prices. We, therefore, also consider the effect of overlapping auctions both on endogenous entry and sniping.

3. Empirical Analysis

Sniping has been proved a rational strategy given common value goods or costly bidding since bidders may avoid price wars by adopting sniping strategy and end up with paying lower prices (Roth and Ockenfels, 2002; Bajari and Hortacsu, 2003). In an absence of these assumptions in real world, whether snipping ends up with higher returns is a question subject to empirical examination.

3.1 Empirical Model and Estimation

As we mentioned earlier, price impact needs to be separated from surplus to make the constitution of returns clear. Thus, we include two sets of regressions in our analysis. Our first set of regressions estimates the price impact of sniping strategy and the second set of regressions estimates the surplus impact of sniping.

Some key considerations underlie our price impact model specification. First, the causal relationship between the sniping strategy and the ending auction price is a complex one, followed by an endogenous issue for the sniping strategy. There is a two way causal relationship between ending price and sniping behavior, referred to as the price impact of sniping and price impact of being a sniper. Logically, it can be that sniping leads to higher (lower) price paid for auctions with certain characteristics that attract snipers who have different value type from the average level. One conventional hypothesis is that snipers are bargain hunters and target auctions with low price level at the last minute before closing. They only snipe the auctions that are worth sniping. That is, types of bidders determined whether they choose to snipe or not. Regardless of the conventional hypothesis, snipers can be associated with high type bidders. If so, although snipers would pay lower than they otherwise would, we cannot necessarily observe that sniping is correlated with lower prices compared to the average. As individual bidders' value type does not correlate with auction ending price, we consider value type possibly an ideal instrumental variable (IV) for the endogenous sniping strategy. We use a bidder's highest bid, which represents his true valuation, to measure the bidder's value type. Another good instrumental variable we consider for the sniping strategy is the buyer's experience. Similar to Bapna et al (2008b), we expect bidder with more experience will demonstrate more strategic bidding behavior such as sniping. Ockenfels and Roth (2002) showed that experts are more likely to bid late. Therefore, buyer experience should be related to the use of snipping, but is not relevant to the price level of the current auction. Generally, a buyer's feedback rating is in proportion to his buying experience. Thus we use buyer feedback and past bidding records to measure buyers' experience.

Except for sniping strategy, we also include the Numbidder, Duration, Sellerfeedback, Crossbidder, Overlap in our explanatory variables for the price impact regression.

According to the past literature, price premium is ascribed to seller reputation (Ba and Pavlou, 2002), which we use Sellerfeedback as a proxy in our regression. Bapna et al (2008b) points out that the price increase brought by seller reputation is weakened by Duration of the auction, a factor we thereby include here to increase explanatory power of the model. Numbidder refers to the number of bidders who place bids for the auction. Number of bidders represents an important aspect of the competition intensity of an auction and is evidenced to be positively associated with higher ending prices (Bajari and Hortacsu, 2003). Theoretically, Bajari and Hortacsu (2003) considered it to be endogenously determined by a zero-profit condition in auction theory. Also, there may be some unobserved auction characteristics that drive up both the number of bidders and the ending price resulting from consumer's high valuation, which constitutes omitted variable bias. To solve this endogenous issue, we use weekend and Startprice as instruments for the number of bidders. If an auction ends in weekend, it will usually attract more bidders (Lucking-Reiley et al, 2007) but it will not necessarily cause auctions to end up with higher prices. Start price is the other instrumental variable we choose. Easeley and Tenorio (2004) showed that early jump bids may impose a threat to the late bidders if bidding is costly and bidders hold private valuation. Thus, higher opening (start price) is supposed to deter bidder entrance. Except for mitigating competition, start price does not directly relate to ending prices.

The Crossbidder and the Overlap between auctions are two variables that measure the competition among auctions due to simultaneity. According to Bapna et al (2009), overlapping auctions share participatory bidders who bid across multiple auctions, which exert a downward pressure on ending prices. We use Crossbidder to measure the number of bidders who bid for similar auctions which overlap with the current auction (crossbid) in time while use Overlap to measure number of auctions that overlap with current auction in time.

To control for the difference between auctions, products and years, we add Reserveflag, Buyitnowflag, product version dummies (halo, premium, elite and arcade) and year dummies (Year2007, Year2008, Year2009). Reserveflag marks whether a seller sets a reserve price for the item while Buyitnowflag marks whether the item has a "buy it now" option or not. We consider both reserve price and "buy it now" option will influence the bidder's valuation and bidding incentive and thus affect the ending price.

To sum up, the set of price equation is composed of three equations and is specified as following:

$$
\begin{aligned}
\ln\_\text{Price} \\
&= \beta_0 + \beta_1\text{Snipe} + \beta_2 \ln\_\text{Sellerfeeback} + \beta_3 \ln\_\text{Numbidder} \\
&+ \beta_4 \ln\_\text{Crossbidder} + \beta_5 \ln\_\text{Overlap} + \beta_6\text{Duration} + \beta_7\text{Reserveflag} \\
&+ \beta_8\text{Buyitnowflag} + \beta_9\text{Year2007} + \beta_{10}\text{Year2008} + \beta_{11}\text{Year2009} \\
&+ \beta_{12}\text{premium} + \beta_{13}\text{elite} + \beta_{14}\text{arcade} + \beta_{15}\text{halo} + \varepsilon \qquad (1)
\end{aligned}
$$

$$
\begin{aligned}
\text{Snipe} \\
&= \alpha_0 + \alpha_1 \ln\_\text{Buyerfeedback} + \alpha_2 \ln\_\text{Pastbidding} + \alpha_3 \ln\_\text{Value} \\
&+ \alpha_4\text{ln\_Sellerfeedback} + \alpha_5\text{ln\_Numbidder} + \alpha_6\text{ln\_Crossbidder} \\
&+ \alpha_7\text{ln\_Overlap} + \alpha_8\text{Duration} + \alpha_9\text{Reserveflag} + \alpha_{10}\text{Buyitnowflag} \\
&+ \alpha_{11}\text{Year2007} + \alpha_{12}\text{Year2008} + \alpha_{13}\text{Year2009} + \alpha_{14}\text{premium} \\
&+ \alpha_{15}\text{elite} + \alpha_{16}\text{arcade} + \alpha_{17}\text{halo} + \varepsilon \qquad (2)
\end{aligned}
$$

$$
\begin{aligned}
\text{ln\_Numbidder} \\
&= \eta_0 + \eta_1\text{Weekend} + \eta_2\text{ln\_Startprice} + \eta_3\text{ln\_Sellerfeedback} \\
&+ \eta_4\text{ln\_Overlap} + \eta_5\text{Duration} + \eta_6\text{Reserveflag} + \eta_7\text{Buyitnowflag} \\
&+ \eta_8\text{Year2007} + \eta_9\text{Year2008} + \eta_{10}\text{Year2009} + \eta_{11}\text{premium} + \eta_{12}\text{elite} \\
&+ \eta_{13}\text{arcade} + \eta_{14}\text{halo} + \varepsilon \qquad (3)
\end{aligned}
$$

We run the estimation by a three stage least square (3SLS) model for two reasons. First, we have endogenous issue for our model. Second, as the error term for each equation comes from the same dataset and may correlate with each other, we probably have dependent regressors. 3SLS combines two-stage least square (2SLS) and seemingly unrelated regression (SUR) and hence takes care of our endogenous issues as well as the potential dependent regressor problem.

The other part of our return story is the surplus regression, in which we replace price with surplus for the dependent variable. The surplus regression model is:

$$\begin{aligned}
\textbf{ln_Surplus} = {} & \boldsymbol{\lambda_0} + \boldsymbol{\lambda_1}\textbf{Snipe} + \boldsymbol{\lambda_2}\,\textbf{ln_Sellerfeedback} + \boldsymbol{\lambda_3}\,\textbf{ln_Numbidder} \\
& + \boldsymbol{\lambda_4}\,\textbf{ln_Crossbidder} + \boldsymbol{\lambda_5}\,\textbf{ln_Overlap} + \boldsymbol{\lambda_6}\textbf{Duration} \\
& + \boldsymbol{\lambda_7}\textbf{Reserveflag} + \boldsymbol{\lambda_8}\textbf{Buyitnowflag} + \boldsymbol{\lambda_9}\textbf{Year2007} \\
& + \boldsymbol{\lambda_{10}}\textbf{Year2008} + \boldsymbol{\lambda_{11}}\textbf{Year2009} + \boldsymbol{\lambda_{12}}\textbf{premium} + \boldsymbol{\lambda_{13}}\textbf{elite} \\
& + \boldsymbol{\lambda_{14}}\textbf{arcade} + \boldsymbol{\lambda_{15}}\textbf{halo} \\
& + \boldsymbol{\varepsilon} \qquad (\mathbf{4})
\end{aligned}$$

$$\begin{aligned}
& \textbf{ln_Numbidder} \\
& = \boldsymbol{\gamma_0} + \boldsymbol{\eta_1}\textbf{Weekend} + \boldsymbol{\eta_2}\textbf{ln_Startprice} \\
& + \boldsymbol{\eta_3}\textbf{ln_Sellerfeedback} + \boldsymbol{\eta_4}\textbf{ln_Overlap} + \boldsymbol{\eta_5}\textbf{Duration} \\
& + \boldsymbol{\eta_6}\textbf{Reserveflag} + \boldsymbol{\eta_7}\textbf{Buyitnowflag} + \boldsymbol{\eta_8}\textbf{Year2007} \\
& + \boldsymbol{\eta_9}\textbf{Year2008} + \boldsymbol{\eta_{10}}\textbf{Year2009} + \boldsymbol{\eta_{11}}\textbf{premium} + \boldsymbol{\eta_{12}}\textbf{elite} \\
& + \boldsymbol{\eta_{13}}\textbf{arcade} + \boldsymbol{\eta_{14}}\textbf{halo} \\
& + \boldsymbol{\varepsilon} \qquad (\mathbf{5})
\end{aligned}$$

3.2 Data

Our analysis is based on a proprietary dataset purchased from eBay, which provides us with full bidding history and transaction information of Xbox 360 consoles auctions from year 2006 to 2010. There are 7318 listed auctions in total. We cleaned the dataset very carefully to make sure that there are no mechanism, product or transaction oriented confounding effects. For example, we filter for the auctions with single unit sale, only dollar as transaction currency, only for new product with no bundles (e.g., with games). We also only consider the auctions that did not sell through "buy it now" option or any other type of alternative buying option. This process resulted in 543 auctions. To take care of the product heterogeneity(there are five types of Xbox 360 consoles, namely, "regular", "premium", "elite", "arcade" and "halo") that we use dummy variables. In the resulting data, the mean ending prices is $449.15 with a standard deviation of 137.78. The median Numbidder for an auction is 7 and the median Startprice for an auction is $225. The average Sellerfeedback and Buyerfeedback are 42.44 and 56.98 respectively, while the median are only 7 and 5, indicating a relatively new market. The mean Duration for an auction is 4.04 days. These statistics are summarized below in Table 1(a).

Table 1(a) Auction Level Descriptive Statistics

	Price	Numbidder	Startprice	Sellerfeedback	Buyerfeedback	Duration
Mean	449.15	7.14	228.89	42.44	56.98	4.04
Median	450	7	225	7	5	3
Q1 (25%)	349	4	50	1	1	3
Q3 (75%)	550	10	390	39	34	7
Stdev	137.78	4.33	176.79	123.74	304.71	2.40
Skewness	0.093	0.65	0.25	6.16	18.97	0.61

Table 1(b) provides some statistics for additional constructs that we use in our study. This includes the information regarding the maximum bid submitted by the bidders and the characteristics of their participation. Value refers to the highest bid that a bidder submits for an auction while Surplus is winner's value minus his final payment which is the second highest price plus a minimum increment. The mean consumer Value for an Xbox 360 console is $474.65

with a standard deviation of 147.45, and the mean Surplus level is $25.5 with a standard deviation of 6.05. Thus, a portion of approximately 5.3% of the valuation goes to consumers' welfare. To obtain the precise measure of bidder's past experience, PastBidding is calculated as the number of all Xbox360 console auctions the bidder has participated before attending the current auction. Since it is a new market, bidder's PastBidding experience is very limited, 0.08 for an average bidder. As mentioned in the introduction, we also consider the price impact of overlapping auctions. We capture the time overlap between auctions by two measures Overlap and CrossBidder. The former one is the number of similar auctions (selling the same type of Xbox 360 console such as both being "premium", which is substitutable for each other). The later one is the number of other auctions overlapping with the current auction. On average, for every current auction, there are 8 other auctions going on. While overlap is prevalent, only a few bidders cross bid, i.e., bid for multiple simultaneously held auctions, which can be seen from the mean value of CrossBidder (mean = 1.14). Due to the skewness of some constructs, in the econometric model we take log function of the Price, Sellerfeedback, NumBidder, CrossBidder, Overlap, Buyerfeedback and Startprice.

Table 1(b) Bidder and Environment Level Descriptive Statistics

	Value	Surplus	Past bidding	Overlap	Crossbidder
Mean	474.65	25.50	0.08	25.97	1.14
Median	485	6.05	0	8	0
Q1 (25%)	356	0	0	4	0
Q3 (75%)	585	25.01	0	45	2
Stdev	147.45	51.76	0.40	30.46	1.87
Skewness	0.08	3.68	6.84	1.15	2.71

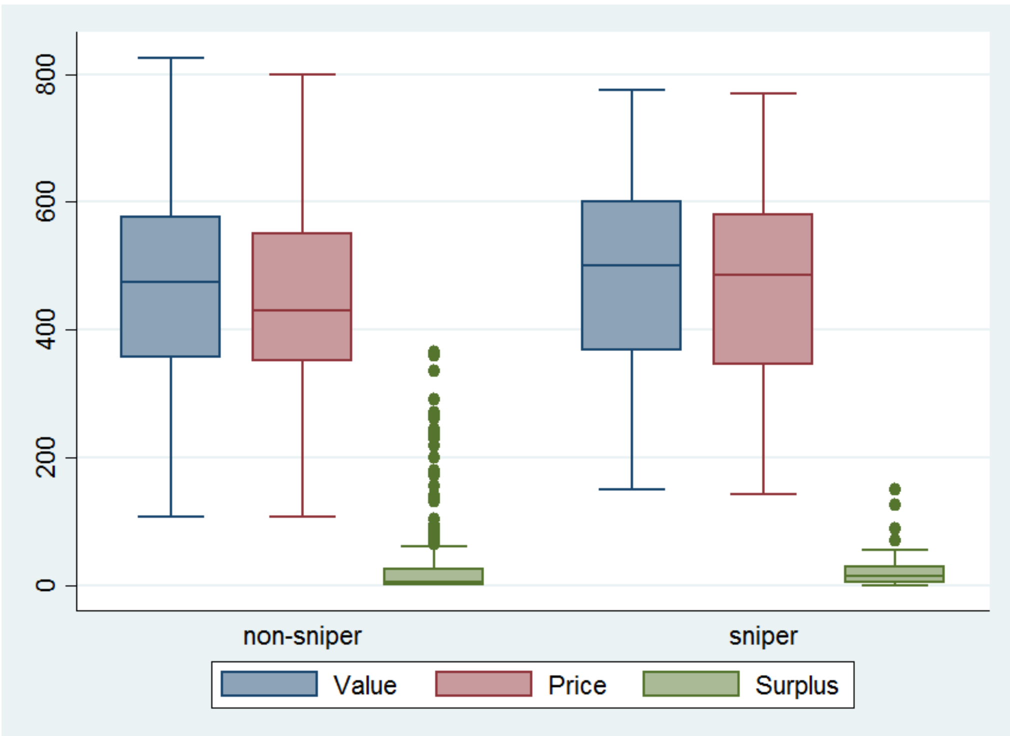
Table 2 presents the correlation table for the dependent variables and main explanatory variables.

Table 2 Correlation between Explanatory Variables

	Price	Snip	Sellerfeedback	Value	Numbidder	Weekend	Crossbidder	Overlap	Duration	Buyerfeedback	Pastbidding	Startprice
Price	1.000											
Snip	0.064	1.000										
Sellerfeedback	-0.162	-0.050	1.000									
Value	0.936	0.050	-0.172	1.000								
Numbidder	0.063	0.104	0.194	0.013	1.000							
Weekend	-0.042	0.051	0.063	0.065	0.012	1.000						
Crossbidder	0.193	0.073	0.047	0.162	0.475	-0.057	1.000					
Overlap	0.544	0.006	-0.126	0.511	-0.037	-0.042	0.257	1.000				
Duration	-0.123	-0.050	0.065	0.089	-0.085	-0.064	-0.110	-0.040	1.000			
Buyerfeedback	-0.013	0.026	-0.011	0.017	0.044	0.082	0.018	-0.012	-0.063	1.000		
Pastbidding	-0.091	0.037	-0.003	0.089	-0.055	0.033	-0.020	-0.023	-0.039	-0.002	1.000	
Startprice	0.356	0.033	-0.170	0.346	-0.655	-0.011	-0.316	0.181	0.001	-0.013	-0.004	1.000

Figure 1 presents a box plot to compare between winners who use sniping strategies and who do not. In figure 1, the left part marked with 0 represents the non-snipers whereas the right part marked with 1 represents snipers. The box plots clearly show that snipers have higher valuation and also pay higher prices than the non-snipers. The surplus level of snipers exhibits lower variance than the non-snipers.

Figure 1 Value, Price and Surplus of Snipers vs. Non-snipers



3.3 Results

Table 3 and Table 4, which we refer to in the following discussion, provide us with an overview of the key estimation results of the price impact regression and surplus regression respectively.

3.3.1 Impact on Auction Price

According to Table 3 (Equations (1) to (3)), we can see that auction ending price paid increases with the use of sniping strategy ($P<0.001$), number of bidders ($P<0.05$), and the number of overlapping auctions ($P<0.001$) while decreases with cross bidding ($P<0.05$) and items with buy_it_now purchasing option ($P<0.05$). This result is consistent with previous literature findings that the number of bidders participating in an auction drives up the ending price. It is also intuitive that buy_it_now price option and cross bidding reduce the auction competition intensity and hence lower the ending price. The new and surprising finding here is that snipers actually pay higher prices. As there are few cross bidders, higher overlap does not necessarily lead to lower competition intensity and lower prices. Besides, bidders with higher valuation is more likely to use sniping ($P<0.001$) while auctions with large number of bidders or overlap highly with other auctions is less likely to attract snipers ($P<0.05$, $P<0.001$). Contrary to the popular belief and description in prior literature that snipers are bargain hunters, our results clearly show that snipers are high value type bidders who target auctions with relatively few bidders and low overlap. That is, they target auctions worth bidding with relatively few substitutes and with price not driven high by the herding effect. Finally, consistent with previous literature, we also see that the number of bidders attracted by the auctions decreases with the start price ($P<0.001$) while positively correlated with sellerfeedback ($P<0.05$). What is interesting in this 3SLS model is the direct reputation effect on price disappears as the correlation between price and seller feedback is not significant. Rather, the reputation increases the number of bidders attracted by an auction and reputation effect is fully mediated by the number of bidders.

Table 3 Estimation Results for Pricing Formation Regression

Method	OLS	3SLS		
DV.	ln_Price	ln_Price	Snipe	ln_Numbidder
	(1)	(2)	(3)	(4)
Snipe	0.031 (0.268)	0.999*** (0.000)		
ln_Sellerfeedback	-0.001 (0.891)	-0.001 (0.956)	0.006 (0.645)	0.035** (0.033)
ln_Numbidder	0.051** (0.007)	0.166** (0.010)	-0.142** (0.015)	
ln_Crossbidder	-0.033* (0.069)	-0.087** (0.045)	0.060 (0.100)	
ln_Overlap	0.118*** (0.000)	0.121*** (0.000)	-0.071*** (0.000)	0.002 (0.930)
Duration	-0.009** (0.025)	-0.004 (0.582)	-0.002 (0.711)	-0.014* (0.101)
Reserveflag	0.043* (0.083)	0.045 (0.274)	-0.037 (0.346)	0.070 (0.199)
Buyitnowflag	-0.046* (0.055)	-0.131** (0.003)	0.115** (0.002)	-0.064 (0.222)
Year2007	-0.169*** (0.000)	-0.099 (0.177)	0.025 (0.711)	-0.054 (0.575)
Year2008	-0.106 (0.202)	-0.393** (0.009)	0.313** (0.017)	0.447** (0.015)
Year2009	0.236*** (0.000)	0.202*** (0.000)	-0.110** (0.008)	0.182** (0.002)
premium	0.169*** (0.000)	0.116** (0.004)	-0.055 (0.153)	0.189*** (0.000)
elite	0.069 (0.329)	0.182 (0.127)	-0.135 (0.222)	-0.013 (0.934)
arcade	-0.753*** (0.000)	-0.839*** (0.000)	0.496*** (0.000)	-0.171 (0.320)
halo	0.413** (0.019)	0.823** (0.007)	-0.694** (0.012)	-0.780** (0.045)
ln_Buyerfeedback			0.014 (0.166)	
ln_Pastbidding			-0.042 (0.559)	
ln_Value			0.618*** (0.000)	
ln_Startprice				-0.161*** (0.000)
Weekend				-0.31 (0.445)

3.3.2 Impact on Bidder Surplus

We present the estimation result for the surplus regression (Equation (4) and (5)) in Table 4. We can see that consumer surplus increases with the use of sniping strategy (P=0.05). This result shows that consumers do benefit from sniping, though the benefits comes more from the value of the products rather than from price cut.

Table 4 Estimation Results for Surplus Regression

Method	OLS	3SLS	
DV.	ln_Surplus	ln_Surplus	ln_Numbidder
	(4)	(5)	(6)
Snip	0.408 (0.061)	0.427** (0.048)	
ln_Sellerfeedback	-0.063 (0.272)	-0.071 (0.229)	0.034** (0.041)
ln_Numbidder	-0.173 (0.229)	-0.079 (0.782)	
ln_Crossbidder	0.055 (0.692)	0.531 (0.770)	
ln_Overlap	0.029 (0.725)	0.029 (0.727)	0.002 (0.936)
Duration	0.077 (0.014)	0.079** (0.010)	-0.014 (0.110)
Reserveflag	-0.036 (0.851)	-0.048 (0.802)	0.069 (0.211)
Buyitnowflag	-0.374 (0.044)	-0.373** (0.041)	-0.066 (0.212)
Year2007	-0.228 (0.496)	-0.212 (0.522)	-0.050 (0.599)
Year2008	0.852 (0.187)	0.802 (0.212)	0.448** (0.014)
Year2009	-0.018 (0.928)	-0.026 (0.896)	0.186*** (0.001)
premium	0.114 (0.539)	0.092 (0.620)	0.187*** (0.000)
elite	-0.867 (0.112)	-0.851 (0.114)	-0.013 (0.935)
arcade	-0.328 (0.582)	-0.343 (0.559)	-0.183 (0.287)
halo	-0.423 (0.757)	-0.360 (0.790)	-0.782** (0.044)
ln_Startprice			-0.165*** (0.000)
Weekend			-0.019 (0.657)

3.3.3 Sub Sample Analysis

To derive further nuance, we tried a sub-sample test using a price impact specification similar to the one for the full sample test but including only the premium, elite and halo version, which totaled 208 observations. Because the market prices for elite, halo, premium and arcade version are respectively \$449.99, \$399.99, \$349.99, and \$279.99. Thus, the elite, halo and the premium version can be viewed as the relatively high stake products compared to the arcade and regular version. The result is shown in Table 5.

From table 5, except for the consistent result that snipers are higher value bidders and pay higher prices, we can see clearly that the use of sniping is positively correlated the bidder's experience measured by buyer feedback ($P<0.1$), and auctions closing on weekend attract fewer bidders ($P<0.05$), indicating a less herding behavior. Previously, List and Lucking-Reiley (2002) found that bidders exhibited more theoretically predicted strategic behavior under high stake items and experienced bidders demonstrate greater tendency to behave strategically. Now our result is quite consistent with past literature and shows that strategic and rational bidding behavior becomes more evident for the high stake sub-sample.

4. Conclusion

In this study, we use a proprietary dataset obtained from eBay which consists of full bidding history for Xbox360 console auctions from 2006 to 2010 to analyze the return of sniping on eBay. Taking care of many observed and unobserved product, seller and auction specific characteristics, we estimate the price impact of sniping in isolation to the price impact of being a sniper. We instrument sniping by the bidder's value type and thus successfully address the

Table 5 Estimation Results for the Sub Sample Test

Method	3SLS		
DV.	ln_Surplus	Snip	ln_Numbidder
	(7)	(8)	(9)
Snip	0.684 (0.000)		
ln_Sellerfeedback	-0.017 (0.234)	0.028 (0.136)	0.043* (0.063)
ln_Numbidder	0.151 (0.085)	-0.182* (0.082)	
ln_Crossbidder	-0.057 (0.252)	0.054 (0.356)	
ln_Overlap	0.070 (0.042)	-0.042 (0.348)	-0.001 (0.977)
Duration	-0.010 (0.218)	0.003 (0.819)	-0.014 (0.307)
Reserveflag	-0.017 (-0.772)	0.058 (0.431)	0.008 (0.934)
Buyitnowflag	-0.048 (0.341)	0.025 (0.710)	0.066 (0.419)
Year2007	-0.103 (0.103)	0.044 (0.595)	-0.096 (0.352)
Year2008	-0.066 (0.647)	0.228 (0.165)	0.506** (0.014)
Year2009	0.335 (0.000)	-0.158** (0.039)	-0.010 (0.913)
premium	-0.253 (0.334)	0.492 (0.120)	1.146** (0.003)
elite	-0.451 (0.050)	0.531 (0.068)	0.933** (0.010)
ln_Buyerfeedback		0.030 (0.086)	
ln_Pastbidding		-0.045 (0.661)	
ln_Value		0.701*** (0.000)	
ln_Startprice			-0.142*** (0.000)
Weekend			-0.122** (0.037)

endogeneity issue in the estimation. From the 3SLS model estimation result, we find that snipers are actually high value type bidders who target premium auctions where the price is not driven higher purely due to herding effect. However, contrary to popular belief, snipers are not bargain hunters, rather they pay higher prices. This result can be also interpreted as the presence of snipers in an auction indicating the premium quality of the auction that leads to higher ending prices. Besides, auctions do not take place in isolation online. Auctions overlapping in time have impact on each other. Given that auctions prevalently overlap with each other, although the pure effect of cross bidding reduces the competition intensity and further lowers the ending price. However, not many bidders take advantage of such cross bidding, leaving room for improvement in buyer welfare and seller efficiency. Last but not least, we establish a causal relationship between seller reputation and auction price, which is different from that in past literature. The positive reputation effect on auction price is fully mediated through the number of bidders attracted by the auction.

This research has laid a foundation for further exploration into allocative efficiency issues in online auctions. As mentioned earlier, only a small fraction of bidders cross bid in overlapping auctions for same items, potentially reducing the allocative efficiency. In future research, we will explore mechanism design issues that can improve the allocative efficiency in overlapping auctions for real world online auctions such as the ones on eBay.

A limitation of this study is that we are unable to isolate the price impact of shilling due to the limitation of the data set. This is an interesting issue that we plan to study in future research via experimental research.

Essay 2 – Measuring the Impact of Crowdsourcing Features on Mobile App User Engagement and Retention: A Randomized Field Experiment

1. Introduction

Currently, mobile app usage accounts for a staggering 15% of the total internet traffic which represents a huge market for app developers and companies. The number of mobile apps grew rapidly from 15,000 in 2009 to over 5.7 million in 20161. Despite a voluminous amount of mobile app downloads2, most apps are notoriously short lived in terms of their user lifetime. According to various mobile app marketing reports, the percentage of apps that is used only once is as high as 20%. Even for apps that are used more than once, over 55% of these users do not use these apps frequently. Moreover, close to 95% of downloaded apps are abandoned by users within a month3. As revealed in Google's 2015 report on mobile app marketing, more than one-third of the users abandon mobile apps due to a loss of interest, indicating that most apps fall short of engaging users for extended periods of time. Moreover, the apps that are abandoned within a short time are found to be the free ones. While the freemium strategy may boost initial downloads, the commitment to use these free apps is extremely low. This suggests that app designers and companies have to look beyond app download volume, and include user engagement and retention as core strategic goals. Taken jointly, the low user engagement levels and high abandonment rate pose great risks for app-based companies and developers, as the profitability of an app is largely dependent on a loyal customer base with regular usage habits. Apps can be monetized in multiple ways, either through a download fee, in-app advertising, or

[1] https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/
[2] The cumulative number of app downloads reached 50 billion for Google Play in 2013, and 100 billion for Apple App Store in 2015, http://www.statista.com/statistics/281106/number-of-android-app-downloads-from-google-play/; http://www.statista.com/statistics/263794/number-of-downloads-from-the-apple-app-store/.
[3] http://whatsnext.nuance.com/customer-experience/95-apps-quickly-abandoned-avoid-becoming-statistic/

in-app purchases. Given that these monetizing strategies can only contribute materially to the developers' bottom line through the sustained usage of their mobile apps, it is imperative for companies to induce user engagement and retention, so as to boost the customer lifetime value of mobile app products.

At the same time, opportunities involving the use of crowdsourcing to enhance user experience and loyalty have risen sharply in recent years. For example, Starbucks (through MyStarbucksIdea) and Dell (through Dell IdeaStorm) have been successfully deployed crowdsourced ideas to refine their product and service offerings. These attempts to involve users as co-creators of product/service design proved to be effective in sustaining user traffic and loyalty to their brand. Moreover, companies such as Wikipedia, Quora, and Threadless (among others) have seen tremendous success by adopting crowdsourcing strategies as their core business operations4. The success of these crowdsourcing-type businesses demonstrates the value in which users place on the crowdsourced content/products that were made available through these platforms.

In light of the issues faced in app usage and opportunities offered by crowdsourcing, we explore new design strategies for enhancing user engagement and retention for mobile apps, by integrating crowdsourcing elements as functionalities within apps. To do so, we first identified two major crowdsourcing processes that can potentially improve user engagement and retention. The process of product co-creation and access to crowdsourced content during product usage are core aspects of crowdsourcing that generates value to the participants of crowdsourcing. The co-creation process endows users with a sense of ownership and control over the design of the final

[4] Wikipedia attracts over 370 million unique visitors monthly as of September 2015, while Quora boasts of a 100 million monthly unique visitors in March 2016. Threadless is estimated to have a $30 million sales and a 30% profit margin in 2012.

product, which can be effective in enhancing their engagement and retention levels. Moreover, products developed with crowdsourced inputs are likely to bear novel features that appeal strongly to the end-users. Thus, the usage of these products would likely lead to greater user engagement and retention.

To investigate the causal impact of crowdsourcing features on user engagement and retention, we collaborated with a mobile gaming company to conduct a randomized field experiment. Through the partnership, we created different versions of the app to reflect the two crowdsourcing sub processes of crowdsourcing and tested the efficacy of these features in improving usage outcomes. Specific to the context of mobile apps, the co-creation process of crowdsourcing translates to the app feature that allows for content contribution, and the product usage process is operationalized as the app feature that enables users to access crowdsourced content. Our experimental results suggest that these crowdsourcing features can be successfully integrated in mobile apps to address the challenge of low engagement and retention rates. Through our experiment, we find that user retention is significantly enhanced by crowdsourcing features, especially when both content contribution and content access features are simultaneously made available to the users. Results suggest that the access to crowdsourced content alone can prove to be helpful in leveraging user retention. We also find that the content contribution feature is largely responsible for increasing user engagement. Moreover, we find user engagement and retention are both enhanced by the mere exposure to the content contribution option. Finally, study results reveal that the impact of crowdsourcing can bear heterogeneous effects on engagement and retention across users with different usage intensity.

Our work makes several contributions. First, by demonstrating how crowdsourcing elements may be used to enhance mobile usage outcomes, our research holds important practical and managerial implications for app companies, designers, and marketers. Given the efficacy of crowdsourcing in arresting low engagement and retention rates, practitioners should consider incorporating crowdsourcing elements within the design of their apps, whenever possible. In this regard, the app designs deployed in our field experiment can provide some guidelines towards how crowdsourcing features may be integrated in mobile apps appropriately. Second, our work contributes to the literature on the antecedents of user engagement and retention in the mobile app context. Despite the extensive literature on user engagement and retention, no studies to date have attempted to examine the impact of crowdsourcing elements on enhancing these key user metrics, as most of the existing work has focused on heightening these usage measures from a marketing perspective. Findings from the focal study provide fresh perspectives on how app usage measures can be improved by harnessing the technological affordance within crowdsourcing which were not attempted previously. In addition to these two contributions, our work makes a third contribution of adding to the literature on crowdsourcing. The extant literature on crowdsourcing has largely focused on examining its impact on product design and its deployment in contests/tournaments, but is largely silent on how a crowd-based strategy might influence usage outcomes. By exploring the finer pathways in which crowdsourcing can affect user engagement and retention via the analysis of detailed tap-stream data, not only does our work contribute to the existing understanding of the broad relationship between crowdsourcing and usage outcomes, it also provides nuanced insights on how individual crowdsourcing features can effectively enhance such outcomes.

2. Literature Review

In our review, we first describe the relevant research on mobile app market as well as the extant crowdsourcing literature. We then draw on literature from several related fields, including human computer interaction, marketing, and information systems to provide a conceptualization of our outcome measures of interest.

2.1. Mobile Apps

Recognizing the need to monetize mobile apps, there is a sizable body of research that examines the impact of pricing strategies on mobile app revenue. For example, Liu et al. (2012) empirically examine the impact of the freemium strategy on Google Play and find a positive association between the use of the freemium strategy and the app's sales and revenue. Through the estimation of consumer demand based on various app characteristics, Ghose and Han (2014) find that app demand varies with the monetization strategy utilized. In particular, monetization via in-app purchases increases demand while monetization via in-app advertising decreases it. In a mobile reading app context, Zhang et al. (2016) find that personalized pricing strategy based on the stage of user engagement can increase total revenue. To our knowledge, there is no existing empirical research that incorporates crowdsourcing features within apps for the purpose of enhancing usage outcomes.

2.2. Crowdsourcing

Crowdsourcing is a digitally enabled strategy that explicitly involves the consumers in the product design process (Howe 2006). Crowdsourcing applications commonly seen in practice include open source software, open content production (e.g., Wikipedia), open innovation contests (e.g., Innocentive), and product co-creation (e.g., Starbucks)[5]. In most settings, the crowdsourcing platform entails collaboration among users to solve problems or

[5] An instance of this is Starbucks utilizing crowdsourced ideas of splash sticks, new coffee flavors, and birthday treats in its operations.

enhance an existing product/service, through which users get to use, evaluate, share, build artifacts, and network with other users (Doan et al. 2011). Broadly speaking, crowdsourcing is composed of two major processes that iterate continuously between each other. The first process is that of product co-creation wherein users provide ideas and inputs for the product of interest, and become directly involved in the product design process. In the purest form of product co-creation, the user gets to specify inputs for her copy of the product and is able to customize the characteristics of her product directly. The crowdsourcing model extends this individual customization model by allowing user inputs to be shared and implemented in a global copy of the product(s) that is used by all users of the community.[6]

Following the product co-creation process, the second process of accessing crowdsourced content during product usage begins. In this step, users other than the contributor try out the new features of the product developed with the crowdsourced inputs. It is through this process that users benefit from products integrating a wide variety of knowledge sources and user tastes for its content. Content contributed by others also make the focal user feel the presence of the user community, especially when the focal user resonate with the crowdsourced content. In a scenario when the focal user can also contribute content, aware of her contributed content being accessed by other users represents an acknowledgement of the contributor's work and is a powerful motivation for users to participate in the co-creation process. Thus, the product usage process is an integral and necessary step for sustaining crowdsourcing models of operation, especially those that operate without financial incentives.

The numerous applications of crowdsourcing in domains pertaining to knowledge

[6] We are aware that such a conceptualization might not apply in certain crowdsourcing contexts in which the crowdsourcer (i.e., initiators in crowdsourcing contests) solicits ideas for internal usage. Hence, we intend to apply our conceptualization strictly for crowdsourcing models that are used on products/services that would made available publicly.

creation have led to a strong focus on understanding the impact of crowdsourcing on idea generation and problem solving. Questions of interest in these studies include assessing the quality of crowdsourced solutions (Jeppesen and Lakhani 2010, Kazai et al. 2011, Poetz and Schreier 2012), evaluating the sustainability of the crowdsourced solution (Bayus 2013, Levine and Prietula 2013, Huang et al. 2014), and uncovering the optimal design of open tournaments (Kittur and Kraut 2008, Boudreau et al. 2011, Wooten and Ulrich 2011, Huang et al. 2012, Wooten and Ulrich 2013).

In a study looking at the quality of crowdsourced ideas, Poetz and Schreier (2012) show that crowdsourced ideas from users tend to outperform ideas generated from professionals in terms of novelty and customer benefit. In the context of open innovation, Jeppesen and Lakhani (2010) find that the effectiveness of crowdsourcing comes from the diversified knowledge of varied sources. Bayus (2013) studied the community of Dell IdeaStorm, and found that individuals are unlikely to consistently submit high quality ideas over time. Boudreau et al. (2011) explore how the number of competitors can influence user participation incentive and solution quality, and Huang et al. (2012) investigate the optimal reward structure for these crowdsourcing contests. Researchers have also examined the effect of feedback on idea generation (Wooten and Ulrich 2011, Wooten and Ulrich 2013), and how collaboration and coordination among crowdsourcing participants can affect group performance and content quality (Kittur and Kraut 2008, Ransbotham and Kane 2011, Ransbotham et al. 2012, Ren et al. 2015). Despite a rich set of studies on crowdsourcing, there are no known attempts to investigate the downstream consequences of deploying crowdsourcing features in product design phase. In particular, the relationships between crowdsourcing and usage outcomes are poorly understood. Furthermore, the mechanisms through which crowdsourcing features affect these usage outcomes remain unknown, which hampers our understanding of how crowdsourcing features may be applied more broadly in the product design process.

2.3. Usage Outcome

User engagement and retention are two core marketing indices that are positively correlated7. However, these two indices refer to different aspects of consumer behavior. While user engagement emphasizes the intensity of user participation, user retention focuses on the longevity of user's loyalty or interest towards a product. Theoretically, the measurements for user engagement and retention can diverge in a scenario where a user demonstrates high initial engagement level but has a relatively short retention due to a rapid build-up of satiation under intensive usage (McAlister 1982, McAlister and Pessemier 1982). Both usage measures are commonly used to assess the customer lifetime value of apps which are helpful in providing profit-related insights.

2.3.1. User Engagement

The user engagement concept originated from two closely related fields, namely human computer interaction (HCI) (Chapman 1997, O'Brien and Toms 2008, O'Brien and Toms 2010) and marketing (Bowden 2009, Vivek 2009, Mollen and Wilson 2010, Verhoef et al. 2010, Brodie et al. 2011, Isaac et al. 2015). Within the HCI literature, user engagement is commonly defined as the interaction between the user and the system (Chapman (1997). O'Brien and Toms (2008) identified interactivity, perceived user control, variety, and novelty as important engagement attributes that characterize the interaction process. The marketing literature views engagement

[7] Webster, J. and J. S. Ahuja (2006). "Enhancing the Design of Web Navigation Systems: The Influence of User Disorientation on Engagement and Performance." MIS Quarterly **30**(3): 661-678.
and Bowden, J. L.-H. (2009). "The process of customer engagement: A conceptual framework." Journal of Marketing Theory and Practice **17**(1): 63-74.
find that user engagement drives customer loyalty and retention in the long run.

as a multi-dimensional concept that goes beyond the mere transactional involvement (Van Doorn et al. 2010), but instead reflects the deeper aspects of consumers' emotional and cognitive connections toward the product (Vivek 2009, Verhoef et al. 2010, Brodie et al. 2011). In the online context, Mollen and Wilson (2010) consider the perceived interactivity to be the precursor to telepresence, which in turn drives engagement.

While common engagement measures include psychological responses measured in consumer surveys (Kim et al. 2013, Isaac et al. 2015), engagement behavior is highly specific to the context in which it is being observed (Brodie et al. 2011, Isaac et al. 2015). This meant that the appropriate measures of user engagement ought to be constructed based on the usage scenarios involved. For example, in studies within educational psychology, user engagement is measured by the time spent on work, the concentration intensity, and the level of contribution of students (Klem and Connell 2004). In the context of online marketing, studies have measured customer engagement through the incidence of activities and actions that are indicative of heightened user involvement, such as the number of posted comments, likes, shares, subscriptions, site click-throughs, bookmarks, emails (Ahuja and Medury 2010); Lee et al. (2015). Literature from HCI and marketing construes usage frequency and intensity as integral parts of engagement, which is reflected by the time spent on using products or organizational offerings (Vivek 2009, O'Brien and Toms 2010). Consistent with this conceptualization, Zhang et al. (2016) has measured user engagement in the mobile app setting via the app usage time. Guided by the conceptualizations in extant literature, we measure user engagement through the duration of users' game session (more details given in a later section).

2.3.2. User Retention

Retention is conceived as a concept that contrasts with switching or churning (Chen and

Hitt 2002). At its core, retention captures the longevity of users' loyalty and interest to a product. Retention is largely measured by the duration of customer's continuous relationship with a service provider (Bolton 1998, Thomas 2001). For example, in Thomas (2001), the retention process begins with the first purchase and continues until the termination of the relationship between customer and the brand. In marketing literature, a major factor that determines customer retention is the extent of variety seeking behavior (Berné et al. 2001), as the desire for change or novelty is related to the satiation that user has from consuming the product (McAlister 1982, McAlister and Pessemier 1982). Van Trijp et al. (1996) find that consumers with low product involvement demonstrate higher levels of variety seeking. In other words, products with low perceived involvement are less likely to retain its consumers. This is consistent with the aforementioned literature that views consumer engagement as an antecedent for customer retention (Bowden 2009).

3. Hypothesis Development

Although many studies have attributed the success of crowdsourcing models to the quality of crowdsourced content, few have looked beyond that to consider how the individual processes within crowdsourcing may contribute to the overall success of crowdsourcing. A critical understanding of the impact of each component within crowdsourcing is a necessary first step towards the appropriate design and integration of crowdsourcing features within products. We discuss the impact of each of these crowdsourcing components and their related impacts on usage outcomes next.

3.1. Product Co-Creation

By turning passive consumers into active co-creators, the process of content co-creation induces users to devote effort-based investment towards the focal product. Consequently, such investments generate user commitment towards the product. Norton et al. (2011) termed this commitment as the "IKEA effect", and reasoned that the investment of effort in providing ideas to enhance the product can lead users to develop an attachment and love for the focal product. In a similar vein, Franke et al. (2010) refer to the "I design it myself effect" to explain why the co-creation process can result in a heightened commitment towards the product. Moreover, being involved in the creation of the final product, users tend to have an elevated valuation of the self-designed product due to a sense of accomplishment and psychological ownership of the product (Pierce et al. 2003). From a task design perspective, co-creation provides users with greater autonomy and perceived control over the focal task and its outcome, which can in turn increase users' commitment to the task (Hackman and Oldham (1976).

The increased levels of attachment and commitment induced by the co-creation process are likely to stimulate one's emotional and cognitive connections to the product, thereby driving user's engagement with the product. A similar relationship has also been proposed by studies in service research, which posit that the involvement of customers in the co-creation process can be an important antecedent for customer engagement (Verhoef et al. (2010). Moreover, by enabling users to provide add-on features or product enhancements through the co-creation process, users would be less inclined to seek out alternatives to the focal product, as they are aware that the product is readily customizable. The autonomy to update product attributes satiates the desire for novelty and variety, which in turn increases the retention rate of users. Accordingly, we propose the following hypotheses:

H1a: Treating users with product co-creation option has a positive impact on user engagement.

H1b: Treating users with product co-creation option has a positive impact on user retention.

The impact of product co-creation on usage outcome could either act through the actual behavior of contributing product-related inputs or through the mere exposure to the option of being able to contribute inputs for the product. According to the co-creation literature, the act of contributing product related inputs directly increases users' sense of ownership and subjective valuation of the product, thereby leading to higher interactivity and usage involvement. Thus, we expect the actual act of product co-creating would increase user engagement. In addition, according to the task design theory, having the option to co-create, would provide users with a greater autonomy over the focal task which increases users' sense of perceived control over the outcome in the long run. Accordingly, we expect that by providing the co-creation option in itself will also enhance the usage outcomes.

H1c: Treated with the product co-creation option, users demonstrate improved usage outcome (i.e., engagement and retention) even without actually participating in the product co-creation.

H1d: With actual participation in the product co-creation, users demonstrate additional improvement in usage outcome (i.e., engagement and retention) to their performance when compared to being merely exposed to the product co-creation option.

3.2. Access to Crowdsourced Content

The second major process within crowdsourcing is access to the crowdsourced content by the community of users during their product usage. Generated from a broad range of knowledge sources, crowdsourced content represents the synergistic integration of ideas across end-users which in turn brings about unprecedented diversity and novelty to the design of the product (Jeppesen and Lakhani 2010, Poetz and Schreier 2012). Hence, products designed with the crowdsourced inputs are likely to bear superior quality and novelty compared to those made by a team of in-house developers (Nishikawa et al. 2013). As such, users are likely to perceive greater variety and novelty in the use of crowdsourced products.

The use of products with crowdsourced content enables users to experience greater product novelty. As such, novel product features created based on user inputs present users with surprising, unfamiliar, or unexpected experiences. Such interactive experiences can be engaging for the user, as novelty appeals to the user's sense of curiosity, encourages inquisitive behaviors, and promotes repeated engagement (O'Brien and Toms 2010). Moreover, products developed with crowdsourced ideas appeals directly to users' interests as these products evolve constantly based on the inputs from other users in the community who contribute novel yet relevant product characteristics. The constantly changing features of the crowdsourced product excite and interest users, thereby reducing their desire to seek variety through alternative substitutes. Accordingly, we propose that:

H2a: Treating users with the option to access crowdsourced content during product usage (abbreviated for Content Access) has a positive impact on user engagement.

H2b: Treating users with the option to access crowdsourced content during product usage (abbreviated for Content Access) has a positive impact on user retention.

Unlike the co-creation process, the effects on content access on user retention are likely to be dependent on the exposure intensity to the crowdsourced content. This is because the emotional and cognitive involvement and the perceptions of novelty cannot be duly experienced if users do not actually see or experience the crowdsourced content itself. Thus, the improvements in usage outcomes are likely to materialize only if users get to see and experience the crowdsourced product, and not through the option of doing so. On a related note, the impact of product usage on usage outcomes would also be dependent on the extent of crowdsourced content that is generated by the user community.

H2c: The actual exposure to crowdsourced content during product usage is needed to have a positive impact on usage outcomes (i.e., engagement and retention).

3.3. Crowdsourcing: Product Co-creation and Access to Crowdsourced Content

Crowdsourcing embodies both of the abovementioned processes by simultaneously allowing users to take part in the product creation process and to access crowdsourced content during product usage. Through crowdsourcing, users interact with other users in the product community implicitly via the product medium through a series of usage-related activities. After a user-contributed feature is released to the community of product users, other users can acknowledge the contributor's efforts by using and testing the product, or by generating further enhancements and features in response to the contributor's work. Through the iterative act of product enhancement (via co-creation), and the usage of different versions of the crowdsourced product, consumers are likely to experience heightened levels of interactivity during their usage of the product, as a result of the implicit communication that takes place on the product artifact itself. The constant product innovation through the incorporation of user-driven enhancements not only raises product quality, but also has the effect of building up a strong community around the product, which introduces a social aspect to the product (Boncheck 2013). Having a social aspect embedded within products has the impact of increasing the value of the product which is shown to raise users' willingness to pay or seek deeper interactions with the product (Oestreicher-Singer and Zalmanson (2012). Such interactions would consequently enhance user engagement.

The joint impact of having the option to co-create and access the crowdsourced content is likely to result in complementary effects that bring about larger effects than the sum of the individual effects. As users learn that their product-related inputs and contributions are accessed or adopted by other users, they develop an attachment to the user community, which can act as a buffer against the tendency to churn and reinforces the willingness to continue product usage. Evidence supporting the positive relationship between community building and user retention has been reported in the marketing literature (Gruen et al. 2000, McAlexander et al. 2002, Algesheimer et al. 2005). Thus, we posit that

H3a: Treating users with the crowdsourcing option has a greater positive impact on user engagement than treating users with either the option of product co-creation or access to crowdsourced content alone.

H3b: Treating users with the crowdsourcing option has a greater positive impact on user retention than treating users with either the option of product co-creation or access to crowdsourced content alone.

4. Experimental Context and Design

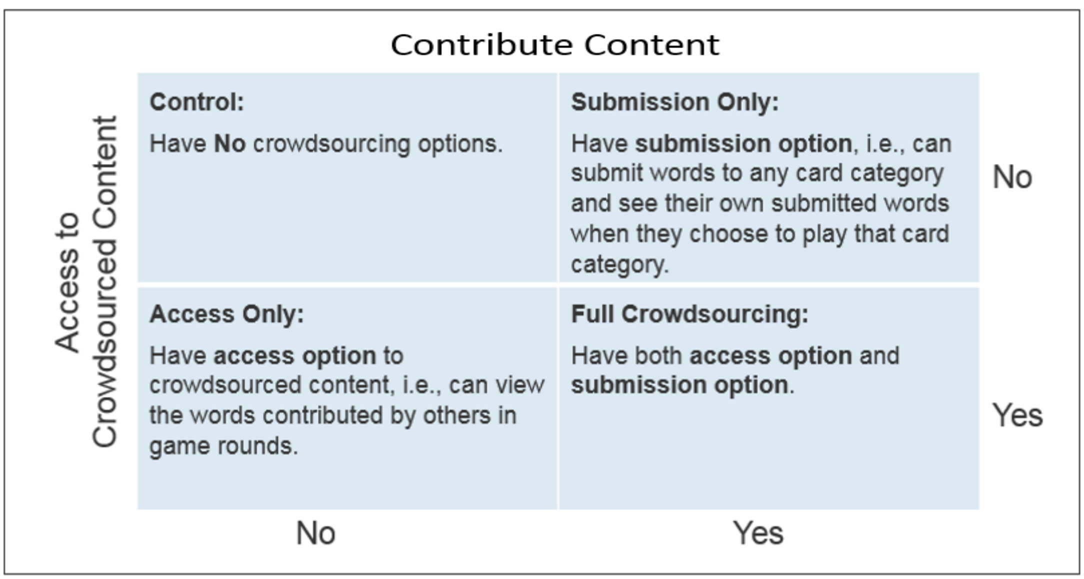
The high abandonment rate of the mobile apps makes it a relevant context for examining the impact of crowdsourcing features on user engagement and retention. In particular, we expect the impact of crowdsourcing to be most pertinent for improving the usage outcomes of content-based apps, especially in gaming apps which require a deal of user participation.

For this purpose, we chose to execute our field experiment via a social mobile game, which is an app version of the popular word-guessing game, "Catchphrase". The gaming app is free to download and monetizes its business through in-app purchases of gaming content. In particular, users are spurred to purchase virtual card decks of in-game words to enhance their gaming experience. Each card deck contains hundreds of phrases that are organized around a particular topic (e.g., "Spring Time", or "Popular Brands"). The phrases contained in card decks constitute the main content of the gaming app and the interestingness of these words would directly appeal to user's engagement with the app. The objective of the game is to let your team mates guess the phrases that show up on the app screen within a time limit, without the use of a certain set of forbidden words. Players organize themselves into two teams, and the team with the most number of correct guesses wins the game. The nature of the game requires its players to be physically co-located in order to engage in a gaming session.

4.1. Experiment Design and Procedures

To test our hypotheses, we partnered with a gaming company to develop different versions of the Catchphrase gaming app. Each version of the app was designed to mirror the various crowdsourcing features discussed above. In particular, we incorporate features that allow users 1) to co-create product content by contributing phrases/words to the card decks and 2) to access crowdsourced content by playing the game with words contributed by the community of gamers. More specifically, we operationalize the co-creation process as a content creation function within the gaming app, and the access to crowdsourced content process as an in-app function that allows user to play the words contributed by other users. Under this operationalization, we utilize a 2×2 factorial setup wherein the control condition is a version of the app that utilizes a word deck that is provided by the app developers, i.e., an app version without crowdsourcing features. The treatment groups are as follows: 1) submission-only group, 2) access-only group, and 3) full crowdsourced group. The submission-only group can submit words to card decks and may view their own submitted words during the game play. However, users in this treatment group are unable to see the words contributed by other users. This is a deliberate design to isolate the effects of crowdsourcing content access from the effects of product co-creation. That said, users in the access-only group can view the words submitted by other players during their game play, but cannot submit words or phrases to the game. The crowdsourced version of the gaming app allows users to both contribute words and view words submitted by themselves and other players during game play. The 2×2 factorial design of our experiment is summarized in Figure 2.

Figure 2 2×2Factorial Experiment Design



To enhance the strength of our experimental manipulation, the app highlights that crowdsourced content during in-game sessions. For instance, in the submission-only and full crowdsourcing version, phrases or words contributed by the focal user bear the banner "Your

Phrase" when they are being played with in a gaming round. Similarly, in the access-only and full crowdsourcing version, words submitted by other users in the community bear a banner "Submitted Phrase" when they appear during the game. Crowdsourced words submitted by users are stored in the card deck together with the words provided by the app developer and are designed to show up randomly in the game. The game developers perform quality checks on all submitted words to remove typos and to ensure that these submissions are appropriately categorized in the appropriate topics.

Furthermore, to ensure that users are able to perceive the word submission feature in the full-crowdsourcing and submission-only versions, the app was designed to have a button labeled "Add Phrases" in a salient location of the screen just before the start of a game session. The purpose of this design is to allow users to perceive the option to co-create product content regardless of whether they actually submitted words to the game. For the full-crowdsourcing and the access-only group, the number of crowdsourced words submitted to a particular card deck is marked on the upper left corner of the card deck (see Appendix A1). Again, this is a deliberate design of placing the crowdsourcing feature in a salient location so that users are aware that their card deck consists of crowdsourced content contributed by fellow gamers from the community.

Prior to the experiment, we estimate the sample size needed to detect the treatment effect. According to the power analysis (Cohen 1988), to detect a small effect size (d = 0.2) with a power of 0.8, under the 0.05 confidence level using a two tail t-test requires a sample size of 394 users for each of the experiment group. To acquire enough users in each group, we conducted the experiment for 113 days such that at the end of the study, we accumulated over 700 users for each control or treated group.

Users are randomly assigned to each of the four groups. The random assignment happens at the instance users install the game app, and users are assigned in a way such that the probability of being each of the four groups is equal. The random assignment of users to the treatment and control group constitute as our main strategy against endogeneity issues that is otherwise present in the analysis of observational data. To eliminate systematic differences from learning and experience, we restrict our experimental sample to first-time users of the app. An additional benefit of relying on a pool of fresh users is that any difference in engagement or retention observed between the control and treated groups would not be attributed to the excitement gained from the new in-game features.

Like all other field experiments, our study faces the challenge of having potential spillover effects between the control and treated groups in the field setting. Although users are not informed of the other app versions besides the one they install, there exists the possibility of users learning about other versions of the app through other users. We run a robustness test in Section 7 to partially rule out the impact of contamination from treatment interference.

4.2. Outcome Variables

To measure users' engagement level, we adopt the definition in O'Brien and Toms (2008), which views engagement as an integrative process of undergoing states of being engaged, disengaged, and reengaged. To measure the level of engagement in our context, we utilize the session duration of each active play session. The session duration is an appropriate measure of user engagement as it captures the usage level of the IT artifact through the time spent on the focal task of interest. This measurement of user engagement in the mobile app context is similar to that utilized by Zhang et al. (2016).[8] Each active session in our gaming app begins from the time the user opens the app to the time she exits the app, which can consist of one or multiple game rounds (See Appendix A1 and A2 for more information on the game setting). In most cases, the act of exiting of the app can be understood as a loss of user interest in the game. As such, a game session in our context represents an instance of continuous game play with which the users expends her energy playing the game till satiation is reached. The more engaged users are, the more game rounds they would play, resulting in longer session durations.

To measure user retention, we look towards the lifetime of the users. Thomas (2001) define the customer lifetime to be the time duration from the customer's first product purchase and till her termination of the relationship with product or service provider. In our game app context, the users churn and do not play the game after some time. We assume a user abandons the game app if a significant amount of time has lapsed since her last game session. Based on established literature on churning (Lewis (2004), the probability of a customer placing an order decreases as the time since the previous order increases. In a similar vein, we expect that the longer the time a user does not log in, the less likely she will return in future. In our data, we find that the duration between two game sessions for 95% of the users falls within 30 days, and the same duration for 90% of the users is 34 days. In other words, the probability of users being active after 34 days of non-activity from their last observed session is very low. Based on the churn literature and the specifics of our context, we define the final day in which the user is active to be 34 days after her last game session. Accordingly, the subject's user lifetime is

[8] Zhang et al. (2016) relies on the amount of time spent reading eBooks in a mobile reading app to measure users' engagement with the reading app.

measured by the number of days between her first and final session.

5. Empirical Analysis

In this section, we first present the summary statistics of the dataset. Next, we present model-free results based on the comparisons of sample means across different treatment groups. Finally, we further explore the finer pathways in which crowdsourcing features drive user engagement and retention.

5.1. Data and Descriptive Statistics

Our field experiment took place over a period of 114 days, during which 3,056 mobile game app users are randomly assigned to one of the experimental groups. The experimental assignment resulted in 757 subjects in the control group, 798 subjects in the full-crowdsourcing group, 742 users in the submission-only group and 759 users in the access-only group. The field experiment yielded 36,938 game rounds spanning over 4,440 game sessions.

We first check the efficacy of the randomization procedure. In particular, we are interested in understanding whether users assigned in each experimental group bears different app downloading behavior or app usage behavior. We do so by looking for differences in download rates by geographical locations and the mobile devices owned by users (proxy by the app download store). We present results of our balance check in Table 6. T-test analyses suggest there are no statistical differences among the control and treated groups with regard to the users' geographic location and their app download sources. We also check for difference in the users' installation patterns over time. The monthly arrival pattern of new users appears to be similar between the control and treated groups, as evidenced in Appendix 4. Taken jointly, it is likely that our assignment of users has produced random samples in each experiment groups.

Table 6. Summary Statistics and Mean Comparison of Full-crowdsourcing, Submission-only, Access-only and Control Group Users

Summary Stats	Control (N = 757)	Full-crowdsourcing (N = 798)	Submission-only (N = 742)	Access-only (N = 759)	t-stats (F – C)	t-stats (S – C)	t-stats (A – C)
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	t-stats (SE)	t-stats (SE)	t-stats (SE)
App Source (Proportion)							
Amazon App Store	50% (0.50)	49% (0.50)	50% (0.50)	52% (0.50)	-0.19 (0.03)	0.13 (0.03)	0.72 (0.03)
Apple App Store	10% (0.30)	10% (0.30)	9% (0.28)	7% (0.26)	-0.27 (0.02)	-0.93 (0.02)	-1.92 (0.01)
Google Play	40% (0.49)	41% (0.49)	41% (0.49)	41% (0.49)	0.36 (0.02)	0.43 (0.03)	0.38 (0.03)
Location (Proportion)							
West-US	17% (0.38)	17% (0.37)	17% (0.38)	17% (0.38)	-0.21 (0.02)	0.04 (0.02)	0.05 (0.02)
MidWest/Central-US	40% (0.49)	40% (0.49)	37% (0.48)	38% (0.48)	-0.01 (0.02)	-1.38 (0.03)	-0.93 (0.03)
East-US	33% (0.47)	34% (0.47)	35% (0.48)	35% (0.48)	0.59 (0.02)	0.99 (0.02)	0.99 (0.02)
Non-US	10% (0.29)	9% (0.27)	11% (0.31)	9% (0.29)	-0.69 (0.01)	0.65 (0.02)	-0.10 (0.02)

Notes: The first 4 columns of this table reports the mean and standard deviation of proportion of user installation from different downloading sources and geographic locations for the control and three treated groups. E.g., the dataset records users who download the game app from Amazon app store as Amazon = 1, otherwise Amazon = 0. The mean value of Amazon equals to 50% for the control group means that 50% of the user installation comes from Amazon app store. Or, the probability that an individual user in the control group install the game app from the Amazon app store equals to 0.5. The last three columns are the results of t-test calculating the mean difference between the control and the treated groups. "F – C" represents the difference between Full-crowdsourcing group and the Control group. "S – C", represents the difference between Submission-only group and the Control group. "A – C" represents the difference between Access-only group and Control group. SD is the abbreviation for Standard Deviation. SE is the abbreviation for Standard Errors. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$.

We further conduct a manipulation test to see if users in the treatment groups perceived the crowdsourcing features and understood how they work.[9] In particular, we check if users in the full-crowdsourcing and submission-only groups understood the word submission feature and were using it in their game-play. As reported in Appendix A5, we find among that 40 out of 798 users in the full crowdsourcing group and 36 out of 757 users in the submission-only group have contributed 171 and 130 words, respectively. While the proportion of users who contributed words are low, such a trend is expected and consistent with the participation inequality rule mentioned by Iyer and Katona (2015) wherein only 1% of the users in the online community would actually contribute content. With approximately 5% of the users contributing words through the full crowdsourcing and submission-only groups, this check suggests that users are indeed aware of the word submission feature in their treatment and are using it in the game.

After assessing the validity of our experimental setup, we next report our main results of the field experiment in Table 7. Based on the kurtosis values of the measures, we note that the key usage metrics are highly skewed. However, the large sample size that we had in the experiment (i.e., over 700 subjects per experimental group) ameliorates concerns related to skewed distributions. Furthermore, a bootstrap simulation shows that the distribution of the sample means conforms to a normal distribution, which meant that t-tests are appropriate for estimating treatment effects based on sample means. By comparing the respective means with the control group, we find that the average session duration is 2.5% higher for the full-crowdsourcing group and 2.61% higher for the access-only group. More importantly, the

[9] By design, we do not conduct a manipulation check for the access feature. To assess whether users perceive the in-game crowdsourced words would involve the intrusive solicitation of user responses. Such an act would make users realize the existence of the experiment, which can introduce unwarranted demand effects. Considering the trade-off of invalidating the field experiment, we decided to mitigate concerns of non-perception by enhancing the saliency of access feature via the app design.

Table 7. Comparison of Full-crowdsourcing, Submission-only, Access-only, and Control Group Users

	Session Duration (in Seconds)					User Lifetime (in Days)				
Experimental Groups	**Mean**	**Median**	**Std. Dev.**	**Skewness**	**Kurtosis**	**Mean**	**Median**	**Std. Dev**	**Skewness**	**Kurtosis**
Control (N = 757)	528.997	120	987.88	4.438	33.436	5.165	1	12.080	3.931	20.142
Full-crowdsourcing (N = 798)	542.46	180	899.78	3.407	20.901	7.769	1	16.976	3.120	12.676
Submission-only (N = 742)	656.291	180	1139.004	3.094	14.402	7.391	1	16.318	3.385	15.343
Access-only (N = 759)	542.800	120	967.229	3.904	25.031	6.771	1	15.850	3.633	17.096

Table 8. T-test Results of Key Play Metrics Difference between Treated and Control Group Users

Play Metrics	**(F – C)**	**(S – C)**	**(A – C)**
	t-stats (SE)	**t-stats (SE)**	**t-stats (SE)**
Session Duration (in Seconds)	0.33 (0.738)	2.75** (0.006)	0.32 (0.746)
User Lifetime (in Days)	3.47** (0.001)	3.01** (0.003)	2.22** (0.027)

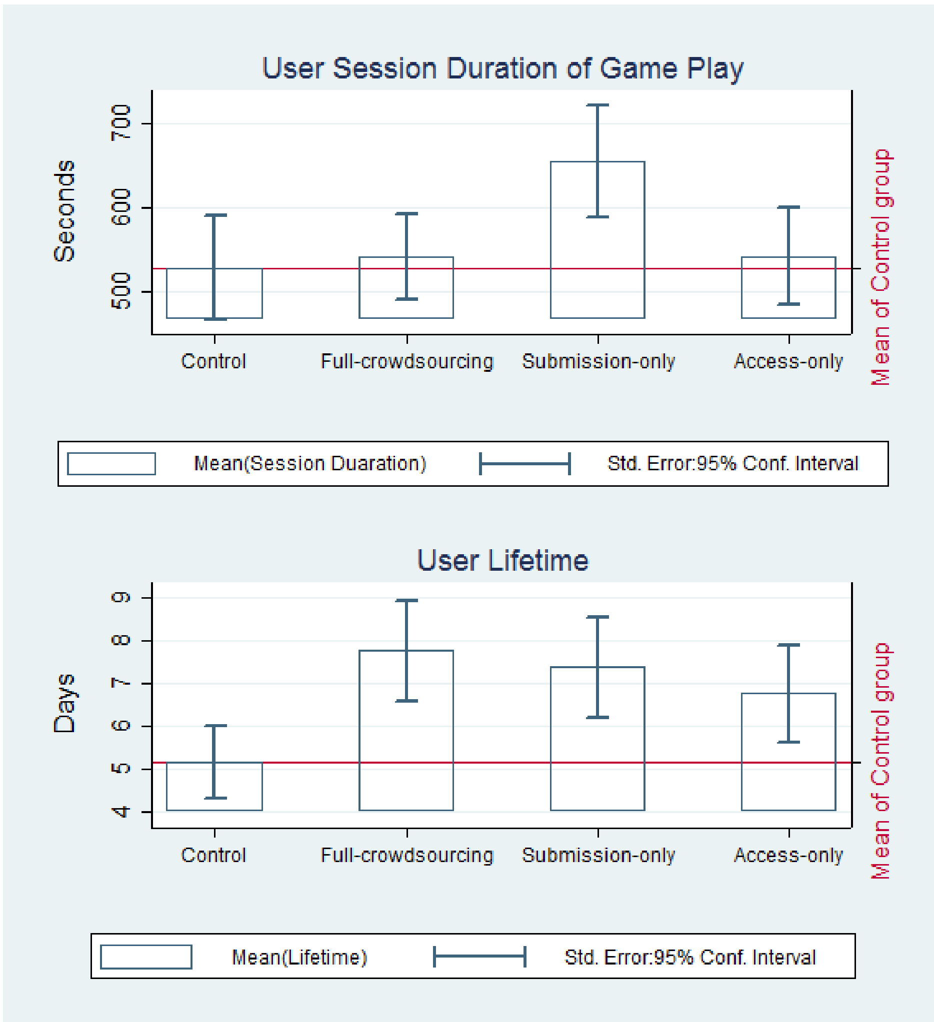
Notes: T-test calculates the mean difference of play metrics between the control and the treated groups. "F – C" represents the difference between Full-crowdsourcing group and the Control group. "S – C", represents the difference between Submission-only group and the Control group. "A – C" represents the difference between Access-only group and Control group submission-only group resulted in longer average session duration than the control group (approximately 24% longer), and is statistically significant (t=2.75, p=0.006). This result supports H1a which states that the co-creation process has a positive impact on user engagement.

Furthermore, we see that the full-crowdsourcing group, submission-only group, and access-only group respectively produced user life that are approximately 50%, 43%, and 31% longer than that of the control group. These differences in mean lifetimes are statistically significant, i.e., the full-crowdsourcing group (t=3.47, p=0.001), the submission-only group (t=3.01, p=0.003) and the access-only group (t=2.22, p=0.027). The results support H1b, H2b, and H3b, which posits that crowdsourcing features have a positive impact on user retention. We summarize these results as error bar charts in Figure 3 for ease of comparison.

5.2. Mechanism Driving User Engagement and Retention

We next examine the underlying manner in which the crowdsourcing features affect user engagement and retention. In particular, we are interested in understanding whether the impact of crowdsourcing on usage outcomes would only act through the actual usage and experience of such features. To do so, we examine and contrast the outcomes of users who participated in actual word submission with outcomes of users who did not submit any word though they were given the option to do so. Similarly, we also look at the differences in outcomes for users who saw crowdsourced words during game play with the outcomes of users who have access to crowdsourced content but did not experience them during game play. To perform this analysis, we used the dummy variable "Word Submission" to indicate whether a user has submitted a word during the current session at time t. The dummy variable "Saw Community Words" is used to indicate whether a user saw words submitted by other users in the player community in the current round at time t. The dummy variable "Saw Own Words" indicates that the user saw her own submitted words in the current round at time t. These binary indicators are added to the regression specification so that the impact of these events can be estimated.

Figure 3 Error Bar Chart of Key Outcome Variables



For this set of analysis, we use the Cox hazard model as our main modeling approach. The Cox model is highly appropriate for our setting, as the model relates the time that passes before the occurrence of an event to covariates that may be associated with that quantity of time. Given that our outcome variables of interest are the amount of time spent in each game session (i.e., session duration) and the time in which users remain active (i.e., user lifetime), the Cox model provides an intuitive way of describing how the hazard of game quitting/churning varies with different crowdsourcing features. Moreover, the Cox model is able to account for the data censoring issues that are common in the analysis of user retention (Bolton 1998). In addition, the Cox model does not require the normality assumption (Cleves 2008) and allows the baseline hazards to vary over time, which further makes it a fitting specification for our purposes.

In particular, we model the hazards of game quitting/churning as a function of a user's treatment and other control variables. The hazard of ending a session or abandoning the game app for user i at time t is specified as follows:

$$\tau(t \mid \delta_i, X_i, Z_{it}) = \tau_0(t)e^{(\alpha\delta_i + \beta \mathrm{X}_i + \gamma Z_{it})}$$

where $\tau_0(t)$ is the baseline hazard. δ_i is a vector of covariates that represent user i's experimental group (full-crowdsourcing, submission-only, access-only, or control), X_i is a vector of covariates indicating user i's time-invariant demographic information, including her geographic location (i.e., West-US, Midwest-Central, East-US, or non-US) and the app download source (i.e., Apple App Store, Google Play, or Amazon App Store), and Z_{it} is a vector of time-varying covariates that affect user i's game play decision at time t. In the estimation involving the hazard of ending a game session, Z_{it} includes the number of players involved in user i's current round, whether the game round is played on a weekend, the tenure of user i in the game, user i's word submission status in the current session and whether crowdsourced word are shown in the current round. For the estimation involving the hazard of abandoning the game app at time t, Z_{it} represents i's word submission status and whether crowdsourced words are shown in the current session. We do not include the number of players as a covariate in the estimation of churning hazard, as player group size is not likely to have an impact on user i beyond the current session. On the other hand, the user lifetime analysis is done at the session level, it is meaningless to aggregate the number of users to be a session level index. In our estimation model, the hazard of ending a session refers to the time between the start and the end of the session, while the hazard of abandoning the game app refers to the time between app installation and abandonment of the game app.

5.2.1. Estimation Results

5.2.1.1. Effects on User Engagement

In Table 9, we report the effects of the submission-only, access-only, and full-crowdsourcing treatments on users' hazard of ending a session. Results show that only the submission-only treatment reduces users' hazard of ending a session by approximately 11% (z=-2.75, p=0.006). After accounting for the various covariates in Model 2, we see that the results are qualitatively similar to the results in Model 1. These set of results affirm the earlier findings that we see in the model-free comparison of group means.

In Models 3 and 4, we explore the underlying mechanisms that drive user engagement by including indicator variables "Word Submission", "Saw Community Words", and "Saw Own Words" in the specification. We compare the submission-only group and the control group in Model 3. Results indicate that users who submitted words in the current session have 37% lower hazard of ending a session compared to users in the control group. Interestingly, even without submitting words to the game, we find that the mere exposure to word submission feature reduces users' hazard of ending a session by approximately 9%. These two estimates are statistically significant. The coefficient of the "Submission-only" variable in Model 3 can be interpreted as the marginal effect of the ability to submit a word, conditional on the actual act of word submission. Hence, these results are in support of H1c. In Model 4, we contrast the subjects in the access-only group with those in the control group. Users who experienced the crowdsourced words generated by other community users do not demonstrate significantly lower hazard of ending a session. Similarly, users exposed to the treatment feature of having access community words in the card deck do not exhibit significantly different level of user engagement. This result does not support H2c.

We provide the results of the same analysis for the full-crowdsourcing condition in Table 10. We find that users who simultaneously submit words and are exposed to the crowdsourced community words have a lower hazard of ending the session compared to control users by 31%, but the significance is only marginal.

Fitted cumulative hazard of ending a session for users in the control group and under different treatment is illustrated in Figure 4, which compare the treated groups with the control group regarding users' cumulative hazard of ending a session. We can see that only the submission-only group, represented by the green curve, demonstrate significantly lower hazard of ending a session than the control group.

Table 9. Cox Proportional Hazard of Ending a Session by Different Treatment Conditions

	Treated Groups vs. Control	Treated Groups vs. Control (Control variables included)	Submission-only vs. Control	Access-only vs. Control
	1 Hazard ratio (p-value)	2 Hazard ratio (p-value)	3 Hazard ratio (p-value)	4 Hazard ratio (p-value)
Treatment				
Full-crowdsourcing	0.966 (0.427)	0.966 (0.415)		
Submission-only	0.887** (0.006)	0.888** (0.03)	0.913** (0.042)	
Access-only	0.977 (0.592)	0.954 (0.282)		0.953 (0.437)
Feature Usage				
Word Submitted			0.627** (0.005)	
Saw Own Words			1.349 (0.529)	
Saw Community Words				1.005 (0.936)
App Source				
Apple App Store		0.791*** (0.000)	0.780*** (0.001)	0.818** (0.012)
Google Play		0.943* (0.069)	0.947 (0.240)	0.995 (0.907)
Geographic Location				
West-US		0.901* (0.097)	0.911 (0.293)	0.883 (0.176)
MidWest-Central		0.843** (0.002)	0.845** (0.031)	0.773** (0.002)
East-US		0.850** (0.005)	0.885 (0.129)	0.797** (0.008)
Number of Players				
5-7 Players		0.846** (0.002)	0.847** (0.032)	0.909 (0.217)
8+ Players		0.942 (0.338)	0.868* (0.098)	0.971 (0.748)
Weekend		0.835*** (0.000)	0.806*** (0.000)	0.850*** (0.000)
User Tenure Age (since installation)		0.997** (0.009)	0.996** (0.005)	0.998 (0.245)
Log Likelihood	-33316.48	-33272.22	-14501.77	-14213.74
χ^2**(df)**	9.05** (3)	97.55*** (12)	74.13*** (12)	38.98*** (11)
Obs (# of Rounds)	36912	36912	18560	16797
Subjects **(# of Sessions)**	4439	4439	2144	2103

Notes: For treatment, the baseline is the control group. For app source, the baseline is the Amazon App Store. For geographic location, the baseline is the non-US users. For number of players involved in a round, the baseline is the option “2-4 Players”. “Subjects” records the number of sessions.

* $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$.

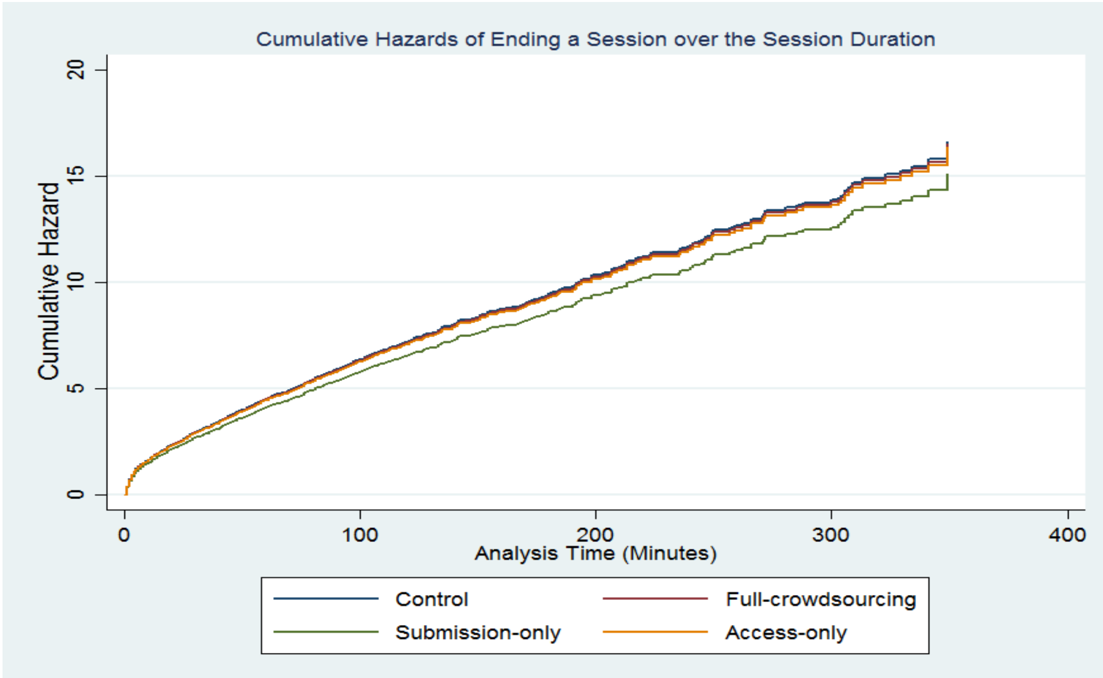
Table 10. Cox Proportional Hazard of Ending a Session in the Full-crowdsourcing Treatment

				Full-crowdsourcing vs. Control
				5 Hazard ratio (p-value)
Treatment				
Full-crowdsourcing				0.951 (0.419)
Crowdsourcing Feature Usage				
	Word Submitted	Saw Community Words	Saw Own Words	
	0	0	1	·
	0	1	0	1.033 (0.608)
	0	1	1	·
	1	0	0	0.745 (0.336)
	1	0	1	·
	1	1	0	0.690* (0.056)
	1	1	1	0.677 (0.698)
App Source				
Apple App Store				0.841** (0.018)
Google Play				0.928 (0.105)
Geographic Location				
West-US				0.943 (0.517)
MidWest-Central				0.877 (0.109)
East-US				0.861* (0.073)
Number of Players				
5-7 Players				0.839** (0.024)
8+ Players				0.957* (0.624)
Weekend				0.767*** (0.000)
User Tenure Age (since installation)				0.999 (0.692)
Log Likelihood				-14978.49
χ^2**(df)**				62.03*** (14)
Obs (# of Rounds)				17531
Subjects (# of Sessions)				2204

Notes: For treatment, the baseline is the control group. For app source, the baseline is the Amazon App Store. For geographic location, the baseline is the non- US users. For number of players involved in a round, the baseline is the option "2-4 Players". "Subjects" records the number of sessions. The missing coefficients denoted by "·" in the table are due to empty or few records for estimation. For example, there cannot be such an instance where the user submitted no words but saw her own words during the game session. Thus, the coefficient estimation is missing for such features usage combination.

* $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$

Figure 4 Treatment Effects on Cumulative Hazards of Ending a Session



5.2.1.2. Effects on User Retention

We repeat the above analysis on the hazard of abandoning the game, and report its results in Table 11. In Models 1 and 2, we present shows the effects of the submission-only, access-only, and full-crowdsourcing treatments on users' hazard of abandoning the game app. Under the basic model without covariates, we see that the full-crowdsourcing treatment reduces users' hazard of abandoning the game app by 17%, the submission-only treatment reduces users' hazard of abandoning the game app by 14%, and the access-only treatment reduce users' hazard of abandoning the game app by 12%. These effects persist with the addition of control variables in Model 2, and are consistent with our earlier results derived from t-tests.

In Model 3, we see that by having the option to submit words is useful in reducing users' hazard of abandoning the game app by 13%. In Model 4, the results suggest that users who saw the community words in the game session demonstrate a 32% lower hazard of abandoning the game app compared to users in the control group. This provides support for H2c through user retention outcomes.

Finally, we report the effect of the full crowdsourcing group in Table 12. Users who were in sessions with community words but do not submit words have a statistically lower hazard of abandoning the app of approximately 30% compared to users in the control group. In addition, users who submit words and see community words in the current session also have a statistically lower hazard of abandoning the game app of 55% in contrast to the control group. The condition where a user submits a word, see her own word, along with other community words within a game session does not produce significant coefficient, possibly due to the small number of such occurrences. Treatment effects on users' hazard of abandoning the game app are illustrated by the cumulative hazards in Figure 5.

Table 11. Cox Proportional Hazard of Abandoning the App by Different Treatment Conditions

	Treated Groups vs. Control	Treated Groups vs. Control (including control variables)	Submission-only vs. Control	Access-only vs. Control
	1 Hazard ratio (p-value)	2 Hazard ratio (p-value)	3 Hazard ratio (p-value)	4 Hazard ratio (p-value)
Treatment				
Full-crowdsourcing	0.830** (0.003)	0.829** (0.003)		
Submission-only	0.860** (0.017)	0.857** (0.015)	0.872** (0.032)	
Access-only	0.878** (0.040)	0.868** (0.026)		1.21 (0.116)
Feature Usage				
Word Submission			0.690 (0.248)	
Saw Custom Words			0.857 (0.765)	
Saw Community Words				0.680** (0.002)
App Source				
Apple App Store		0.680*** (0.000)	0.634*** (0.000)	0.658** (0.001)
Google Play		0.925* (0.102)	0.888* (0.080)	0.857** (0.023)
Geographic Location				
West-US		0.810** (0.019)	0.907 (0.437)	0.733** (0.014)
MidWest-Central		0.859* (0.055)	0.872 (0.221)	0.828* (0.088)
East-US		0.778** (0.002)	0.776** (0.030)	0.767** (0.022)
Log Likelihood	-14369.89	-14353.40	-6411.52	-6415.79
χ^2(df)	9.86** (3)	42.83*** (8)	30.40*** (8)	33.85*** (7)
Obs (# of Sessions)	4440	4440	4440	4440
Subjects (# of Users)	3056	3056	3056	3056

Notes: For treatment, the baseline is the control group. For app source, the baseline is the Amazon App Store. For geographic location, the baseline is the non-US users. "Subjects" records the number of users. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$

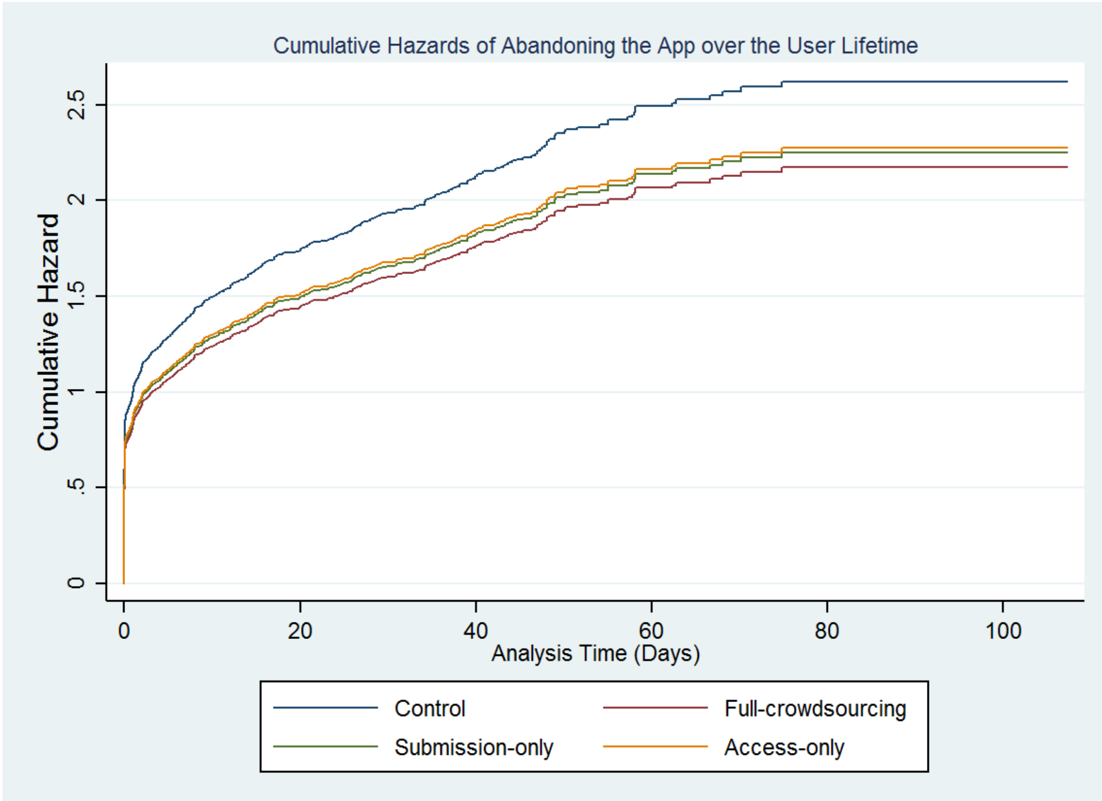
Table 12. Cox Proportional Hazard of Abandoning the App in the Full-crowdsourcing Treatment

			Full-crowdsourcing vs. Control
			5 Hazard ratio (SE)
Treatment			
Full-crowdsourcing			1.145 (0.296)
Submission-only			
Access-only			
Crowdsourcing Feature Usage			
Word Submission	Saw Community Words	Saw Custom Words	
0	0	1	·
0	1	0	0.706** (0.008)
0	1	1	·
1	0	0	0.000 (1.000)
1	0	1	·
1	1	0	0.452** (0.025)
1	1	1	0.477 (0.151)
App Source			
Apple App Store			0.729** (0.006)
Google Play			0.946 (0.405)
Geographic Location			
West-US			0.809* (0.098)
MidWest-Central			0.836 (0.111)
East-US			0.761** (0.020)
Log Likelihood			-6681.37
χ^2(df)			33.37*** (10)
Obs (# of Sessions)			2204
Subjects (# of Users)			1555

Notes: For treatment, the baseline is the control group. For app source, the baseline is the Amazon App Store. For geographic location, the baseline is the non-US users. "Subjects" records the number of sessions. The missing coefficients denoted by "·" in the table is due to empty or few records for estimation. For example, there cannot be such an instance where the user submitted no words over her user lifetime but saw her own words. Thus, the coefficient estimation is missing for such features usage combination.

* $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$

Figure 5. Treatment Effects on Cumulative Hazards of Abandoning the Game App



5.3. Heterogeneous Effects

It is plausible that the crowdsourcing features may have differential impact across different users. Some users may exhibit stronger responses to these features by extending their game sessions or lifetime while others do not. The large dispersion of the outcome variables (see Table 7) suggest that the possibility of heterogeneous effects across users.

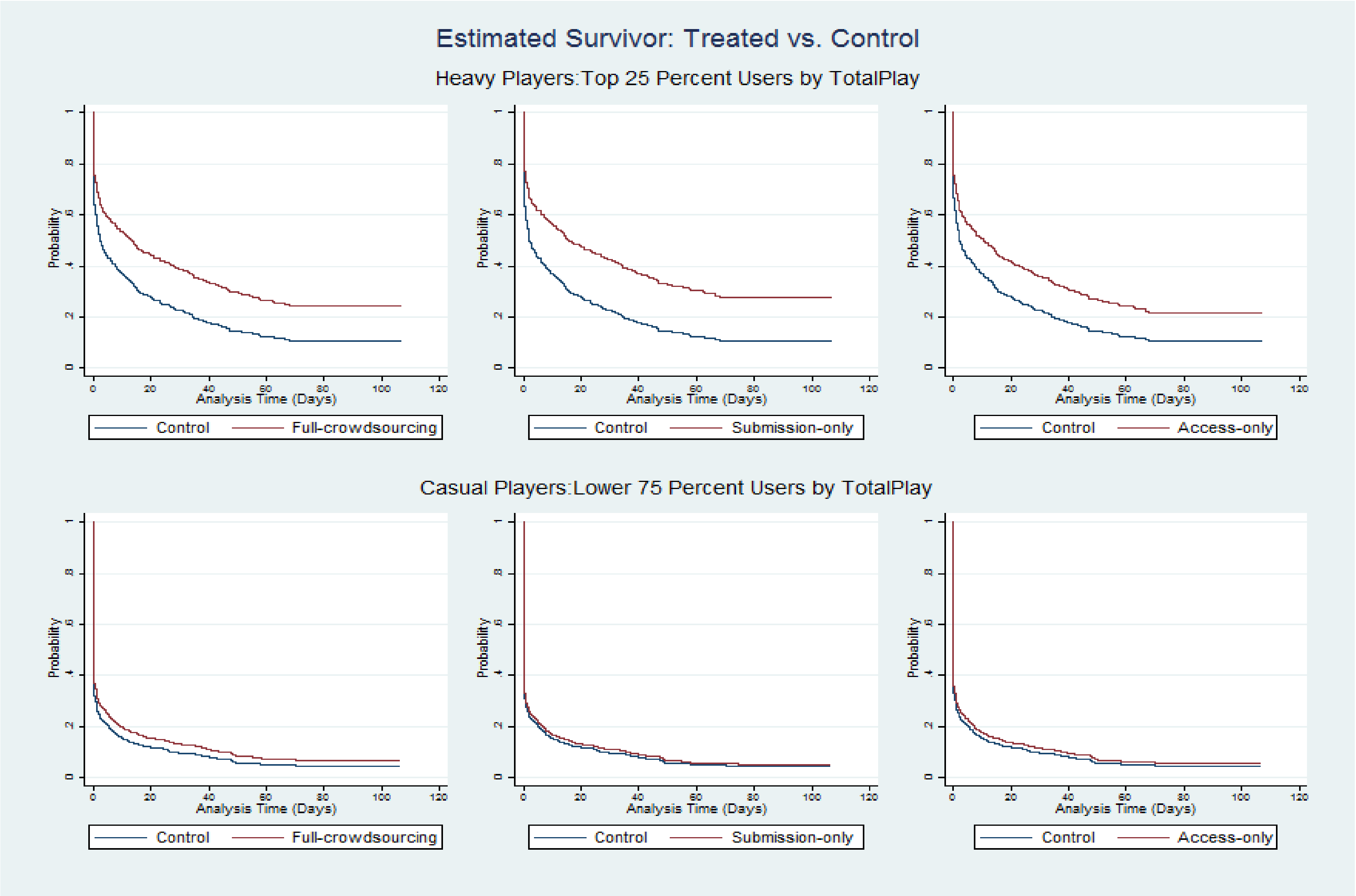
In our mobile game app context, heavy users are likely to exhibit different behaviors compared to casual users. For instance, the heavy users are likely to be endowed with higher involvement levels which results in longer gaming time. Highly involved users are likely to react to crowdsourcing features more favorable than the light users, which meant these features would likely have salient greater impact on these users in terms of their usage outcomes.

To test for this possibility, we split each of the experimental groups into two segments based on the users total game play duration. Users from the top 25 percentile from each group represent heavy players while the remaining users are labeled as casual players. We next perform separate regressions on each of the user segments. Results from Table B1.1 (Appendix B) show that the various treatments do indeed hold different types of impacts on heavy users and casual users. In particular, crowdsourcing features tend to be more effective in enhancing the engagement levels of casual users compared to heavy users. However, the opposite trend is true for retention outcomes: crowdsourcing is more effective in enhancing the retention of heavy users compared to casual users.

From Models 1 and 2, we see that the casual users in full-crowdsourcing group and submission-only group had a significantly lower hazard to end their game session (17% and 16%, respectively). However, the engagement levels of heavy users do not seem to be affected by the crowdsourcing features, except for those in the submission-only treatment group. In particular, we note that the full-crowdsourcing treatment does not seem to have a significant positive impact on user engagement. From Models 3 and 4, we see that for the heavy users' hazard to abandon the game app is significantly lower in the full-crowdsourcing treatment (37%), the submission-only treatment (42%) and the access-only treatment (32%), compared to the control group. On the other hand, only the full-crowdsourcing treatment reduce casual users' hazard to abandon the app by 13%. The difference of treatment effect over user lifetime between the casual users and the heavy users is illustrated in Figure 6. By comparing the difference of survival rate between the treated groups and the control group, we see that the treatment effect on user lifetime is much greater for the heavy users than the casual users.

This set of results might arise due to heavy and light users valuing in-game autonomy and content novelty differently. On one hand, heavy users may value product content variety more than casual users. According to McAlister (1982), variety seeking is more likely to happen when consumption of certain product attribute accumulates to reach the satiation point along users' consumption history. This implies that, the heavy users in our experiment who consumed more of the game content than casual users require higher level of content variety or novelty to sustain their continued interest in the game. Thus, crowdsourcing features, which bring more novelty to the game, will have more salient impact on user retention on heavy users. On the other hand, compared to casual users, heavy users weigh autonomy even more than product content variety such that the full-crowdsourcing group is outperformed by the submission-only group in terms of user engagement level.

Figure 6 Survival Probability over User Lifetime of Different User Segments



6. Robustness Check

Next, we assess the robustness of our results. First, we substitute session duration with an alternative dependent variable that reflects total play duration, to assess if our findings on user engagement are sensitive toward operational definitions. Second, we execute a sensitivity analysis to assess whether the treatment effect on users' hazard of abandoning the app hold under different lifetime censoring thresholds. Third, to partially exclude the impact of possible contamination, we include in our regression the interaction terms of treatment with geographic location or number of players. Finally, we examine alternative survival analysis models to see if these models provide a better fit with the data.

To corroborate that our findings on user engagement is not sensitive to the operationalization choice of the outcome variable, we rely on total play duration as an alternative outcome variable for user engagement. This alternative measure is captured by the total time spent on the game app within the user's lifetime, which is calculated by the aggregation of the session durations over users' multiple log in accesses to the app. The total play duration as an outcome variable has the advantage of integrating users' short term interest and play intensity, access frequency and endurance of game play into a single compound measurement. The analysis result is shown in Table B2.1 (Appendix B2). The results show that both the full-crowdsourcing feature and submission-only feature have a significant positive impact on enhancing users' sustained engagement, while the access-only feature does not have a significant impact on users' total time spent on the game. Compared to measuring short term engagement level via session duration, the effects of the full-crowdsourcing feature on users' sustained engagement is now significant, possibly due to that the compound nature of outcome variable.

We also checked whether using 34 days is an appropriate cutoff point to denote the user churn behavior. We tested our regression model with threshold values of 30 days, 38 days and 46 days, which ranges from the 89 percentile to the 95 percentile among all users. The regression results remain qualitatively similar as our main regression results (See Table B2.2 in Appendix B2). Although users are randomly assigned to different treatment groups, it is nevertheless difficult in preventing subjects from different experimental groups from speaking to one another, which can lead to a spillover of treatment effects. In general, geographic locations with higher user density are more likely to suffer from contamination issues, as the proximity of individuals can lead to a greater likelihood of communications taking place between users across different experimental groups. Under this possibility, locations with higher user density would exhibit systematic differences in treatment effects from locations with low user density. To assess the possibility of contamination, we first interact the treatment group with geographic location, and include these interaction terms in our regression model. If users were contaminated in the prescribed fashion, we will find a significant interaction effect between treatment and geographical location. A similar check can be performed using the interaction between treatment group and number of players, as co-players is likely to download and use the game app in future. Given that the higher number of players increases the chances of contamination, possible contamination due to comparisons among gaming peers may result in a different usage outcomes especially in game sessions that are played with large number of players. If true, the interaction term between the treatment group and the number of players would produce significant estimates. According to our regression results in Table B2.3 (Appendix B), neither the interaction between treatments and geographic location, nor that between treatments and number of players has a significant impact on users' hazard of ending a session. Similarly, Table B2.4 shows the interaction terms do not have a significant impact on users' hazard of abandoning the app. Thus, it is likely that the experiment does not suffer from contamination due to high user density or large number of co-players in a game.

Finally, we replace our Cox models with different classes of survival models to see if alternative modeling approaches might provide better fit to our data. In this exercise, we attempt the use of the Proportional Hazard parametric models (PH models) assuming exponential, Weibull, and Gompertz distribution for the failure times. Results from these PH models are qualitatively similar to our results derived from the Cox model. We then make estimations using the Accelerated Failure-time models (AFT models) which assume generalized gamma distribution for failure events. The estimated κ and σ parameter indicate that none of the nested models with exponential distribution, Weibull distribution or the lognormal distribution provide a good fit with our data. We compare the Cox-Snell residuals and found that the overall goodness of fit of our Cox Proportional Hazard model is better than the generalized gamma regression. In particular, the cumulative hazard of Cox-Snell residuals estimated by our Cox Hazard model for the hazard of ending a session lies perfectly along the 45 degree reference line (See Figure B2.1 in Appendix B2).

7. Discussion

In this work, we attempt to examine how the different processes of crowdsourcing, product co-creation and product usage, might be implemented as in-app features for boosting user engagement and retention. We designed a field experiment under the context of a mobile gaming app to test our hypotheses. In our field experiment, we operationalized product co-creation as a feature that allows users to create and submit in-game content, and product usage as a feature that allows users to access and play with the in-game content created by other users. The experiment revealed several interesting insights on how each of these crowdsourcing features affect usage outcomes. We find that user engagement can be enhanced by the content submission feature, and that user retention is enhanced by both the individual and combined features of content submission and content access. Results also show that the actual act of submitting words along with the mere exposure to the word submission option are effective in increasing user engagement, while the actual experience of playing with crowdsourced words in the game (not the awareness of having access to the crowdsourced content) is necessary in driving user retention. We further find that heavy users tend to be affected by each of the crowdsourcing features differently from the casual user. Finally, we note that the study results are robust towards alternative variable operationalization and model specification. We next discuss some of the theoretical contributions and practical implications of our work.

7.1. Theoretical Contributions

While there has been a good amount of research on the topic of crowdsourcing, there is still a limited understanding of how the various processes embedded within crowdsourcing might be harnessed as product features to enhance usage outcomes. Our work contributes to the literature of crowdsourcing by assessing and quantifying the causal impacts of its sub-processes on user engagement and retention. Our work finds that each of these crowdsourcing processes can have varying impacts on different usage outcomes. In particular, user engagement is only enhanced by product co-creation and not access to crowdsourced content option. Such a result suggests that the content creation aspect of crowdsourcing increases user engagement by giving users perceived control over the focal task, which consequently raises their involvement and interest for the app.

Second, results show that when both features of word submission and crowdsourced content access are provided simultaneously, users who submit words and see crowdsourced words experience the greatest increase in user enhancement. Users who submit words but do not see crowdsourced words in return show no significant improvement. At a first glance, it's surprising that the full-crowdsourcing features provide no significant enhancement after mediating the treatment effect through the actual feature usage, whereas the mere existence of its sub option, the word submission feature, significantly leverages user engagement in the submission-only treatment. The significant enhancement brought by the mere existence of the word submission option in the submission-only treatment can be explained by users' sense of control over the game app. The users feel more empowered in customizing the game app in the submission-only version since all crowdsourced words randomly drawn into the game round are from the user herself. On the other hand, the full-crowdsourced version users have little manipulation over the words shown in their game round through submitting words since the crowdsourced words are drawn from a much larger pool including the words submitted by other users. According to our data, the probability to see custom words for users who submit words in the full-crowdsourcing group is only 1/3 of that for users who submit words in the submission-only group[10]. Thus, it is the different level of control between the full-crowdsourcing treatment and submission-only treatment that leads to the different impact on user engagement by the mere existence of the full and partial crowdsourcing features.

The results also suggest that both the word submission option and the access to crowdsourced words option leverage user retention. While the mere existence of the word submission option leverage user retention in the submission-only treatment, users retention gets significantly improved only when they are actually exposed to crowdsourced words in the

[5] The user has submitted word in that round, controlling for the round duration. Please see Appendix B3 for the regression result.

access-only treatment. Thus, users' interest in the game app does get sustained by the more diversified content that evolves over time. Besides, when users are offered the full-crowdsourcing feature, greatest improvement in user retention is achieved when users both submit words and see crowdsourced words in a session.

7.2. Managerial Implications

Due to our peculiar research setting, we gain a perspective of customization in the framework of crowdsourcing. The managerial implication is when companies consider about the co-creation strategy, they should take users' preference product control into consideration to determine whether to use a customization strategy or the co-creation strategy. Especially for the heavy users, it would possibly be better to keep a customization version aside from the full-crowdsourcing option.

Past crowdsourcing literature ascribe the merit of crowdsourcing to be the supreme solution quality, while the role of crowdsourcing as a content sharing and community building tool to enhance user retention has been largely downplayed.

8. Conclusion

With all our findings discussed above, we consider our contribution includes the following aspects. First, we are the first to measure the causal impact of crowdsourcing features on user engagement and retention for the mobile app context. Second, we explore the underlying mechanism through which crowdsourcing features drive user engagement and prolong user retention. Due to our peculiar research setting, we gain a perspective of customization in the framework of crowdsourcing. The managerial implication is when companies consider about the co-creation strategy, they should take users' preference product control into consideration to determine whether to use a customization strategy or the co-creation strategy. Especially for the heavy users, it would possibly be better to keep a customization version aside from the fully interacted crowdsourcing option. Besides, since the effectiveness of employing content access option to enhance user retention depends largely on users' actual exposure to the crowdsourced content, the success of crowdsourcing strategy largely depends on whether the company is able to solicit enough content from the crowd. Finally, the past crowdsourcing literature ascribe the merit of crowdsourcing to be the supreme solution quality, while the role of crowdsourcing as a content sharing and community building tool to enhance user retention has been largely downplayed.

In this paper, we also test a novel approach for sustaining mobile app usage from a product design aspect, which is distinguished from the pricing strategy given in the past literature.

Conclusion

For the better design of online platform that maximizes the transaction or usage efficiency, it is quite important to find out how users react to different online settings and how platform efficiency is involved.

In my first essay, I investigate the efficiency of the eBay online auction platform by examining its bidder welfare. The research question of how bidders' sniping strategy, which is enabled by eBay's peculiar proxy bidding setting, affects bidder surplus has been pursued. Though sniping is believed to lead to higher buyer efficiency by the conventional wisdom, empirical evidence is missing from the extant literature due to the following data challenges. i) sniping is endogenous and its causal impact on bidder welfare is difficult to quantify using observational data, ii) the highest bid of an auction is not revealed to the public due to the proxy bidding setting on eBay and researchers thus lack the bidder valuation to calculate winner surplus. Using a unique dataset that contains the full bidding history for Xbox 360 consoles on eBay over a period of 5 years, I was able to examine the impact of sniping strategy on auction prices and bidder surplus. More specifically, since eBay auctions are an approximation of second price sealed-bid auction mechanism, the winner's highest bid can be viewed as their true valuation of the auctioned item (Vickery 1962). I thus measure winner surplus by the difference between winner's highest bid and the price paid. When estimating the causal impact of sniping on auction ending price, my focus is on isolating the price impact of sniping from the price impact of being a sniper. Since sniping is likely to be endogenous to bidder's value type while auction ending price is independent of the winner's valuation due to the truth-telling bidding strategy, I use the bidder's valuation as the instrumental variable for sniping. The key results indicate that contrary to popular belief, snipers may not be bargain hunters. Rather, they are value shoppers and often pay higher prices as compared to non-snipers. But overall, bidders do benefit from snipping as they get higher surplus. The results also indicate that higher prices due to higher reputation, a well-established result in the literature, are mediated by attracting snipers and/or higher number of bidders in an auction. Finally, I find evidence that bidders' strategic behavior becomes more evident at a higher stake value. These findings show that the sniping strategy not only improves the bidder welfare by increasing their winning surplus, but also the increased auction ending price is indicative of increased seller welfare. Thus, the use of sniping strategy by bidders actually increases the total efficiency of the eBay auction platform. Our study here will be helpful to auction platform owners for their understanding of bidder segments. And based on the understanding, platform settings can be improved to better match bidders and auction items. For example, the auction platform owners can think about how to direct snipers to quality items that deserve high price through certain platform setting. By allocating high value bidders to high quality items, the auction platform will further improve their efficiency.

In the second essay, I explore the efficacy of crowdsourcing features on enhancing user engagement and retention in the context of mobile gaming apps. To do so, I examine two specific crowdsourcing features, namely, the ability to contribute content and the ability to access crowdsourced content. Under a 2×2 factorial design, I assess the impact of these crowdsourcing features on usage outcomes via a randomized field experiment. In our experiment, I also examine the underlying mechanisms that lead to the main observed effects. Interestingly, I find that even without contributing content, users exhibited heightened user engagement and retention when they are given the option to contribute gaming content to the app. Results also show that the positive effect of content access on user retention only materializes when users are actually exposed to crowdsourced content. In particular, results show that the content contribution feature reduces users' hazard of ending a session and abandoning the app by 11% and 14% respectively. Moreover, the largest improvement in user retention is achieved when users are able to contribute content and view crowdsourced content from the community. Finally, I find heterogeneous treatment effects in our study setting: crowdsourcing features tend to enhance the retention of heavy game players, while they heighten the engagement level of casual gamers. In this paper, I fill the gap of the crowdsourcing literature by drawing from marketing, organization science and human computer interactive literature to theorize about how crowdsourcing features can improve usage outcome in terms of engagement and retention level. I then complete the study by empirically testing our theory using a randomized field experiment. The empirical results proved that crowdsourcing features enhance both user engagement and retention as is proposed in our theory part. Also, our empirical findings on the mechanism through which crowdsourcing affects usage outcomes provide practical guidance for merchants. Merchants who want to improve usage outcome by incorporating crowdsourcing features into their products can refer to our findings for whether to provide a customized version or a full crowdsourcing version to a particular user segment.

As a final conclusion, in this dissertation I studied two digital enabled platforms, an online auction site and a social mobile game app, for how their efficiencies are affected by user behavior under certain platform settings. My dissertation provides both theoretical and empirical insights into how sniping strategy enabled by the proxy bidding setting and crowdsourcing features of the game app respectively lead to higher platform efficiency. My study is subject to some limitations due to data availability. In our first essay, we are not able to distinguish shilling bids, which is used by sellers to protect themselves from low ending price, from the sniping bids in our data. Thus, our result is in certain extent vulnerable to the possibility of shilling. However, the identification of shilling bids and isolating the impact of shilling from the surplus estimation for true snipers can be a future topic for empirical research. On the other hand, our crowdsourcing study can be extended in the future by estimating the usage impact of other options that enrich the crowdsourcing feature, such as the option to allow users to submit ratings or feedback for the crowdsourced content.

Bibliography

Ba, S., Pavlou, P.A., (2002), "Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior", *MIS Quarterly*, 26(3): 269-289.

Bajari , P., Hortacsu., (2003), “The winner’s curse, reserve prices, and endogenous entry: Empirical insights from eBay auctions”, *The Rand Journal of Economics*, Vol.34, No.2, pp 329-355.

Bapna, R., Chang, S.A., Goes, P., Gupta, A., (2009), "Overlapping online auctions: Empirical characterization of bidder strategies and auction prices", *MIS Quarterly*, (33: 4) pp.763-783.

Bapna, R., Jank W., Shmueli, G., (a), (2008), "Consumer surplus in online auctions", *Information Systems Research*, 19:(4), pp 400-416.

Bapna, R., Jank W., Shmueli, G., (b), (2008), “Price formation and its dynamics in online auctions”, *Decision Support Systems*, Vol. 44, No.3, 2008, pp 641-656.

Dellarocas, C., (2003), “The digitization of word-of-mouth: Promise and challenges of online reputation systems,” *Management Science* 49:(10), pp 1407-1424.

Easeley, R.F., Tenorio, R., (2004), “Jump bidding strategies in internet auctions”, *Management Science*, Vol. 50, No.10, pp 1407-1419.

Gray, S., Reiley, D., (2007), "Measuring the benefits to sniping on eBay: Evidence from a field experiment,” working paper available at *http://www.u.arizona.edu/~dreiley/papers/Sniping.pdf*.

Hu, W., Alvaro, B., (2008), “Online auctions efficiency: A survey of eBay auctions”, Alternate track: *Industrial Practice and Experience*, April 21-25, Beijing, China.

List, J.A., Lucking-Reiley, D., (2002), "Bidding behavior and decision costs in field experiments," *Economic Inquiry*, Oxford University Press, vol. 40(4), pages 611-619.

Lucking-Reiley, D., Bryan, D., Prasad, N., Reeves., D., (2007), “Pennies from eBay: The determinants of price in online auctions”, *The Journal of Industrial Economics*, Vol. 55, No. 2, pp. 223-233

Milgrom, P.R. and Weber, R.J., (1982), “A theory of auctions and competitive bidding”, *Econometrica*, Vol. 50, pp. 1089-1122.

Roth, Alvin E. and Axel Ockenfels, (2002), "Last-minute bidding and the rules for ending second-price auctions: Evidence from eBay and Amazon auctions on the Internet," *American Economic Review*, 92 (4), 1093-1103.

Vickrey, W., (1962), “Auctions and bidding games”, *Recent Advances in Game Theory,* 29 (1962): 15-27.

Ahuja, V. and Y. Medury (2010). "Corporate blogs as e-CRM tools–Building consumer engagement through content management." *Journal of Database Marketing & Customer Strategy Management* 17(2): 91-105.

Algesheimer, R., et al. (2005). "The social influence of brand community: Evidence from European car clubs." *Journal of Marketing* 69(3): 19-34.

Bayus, B. L. (2013). "Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community." *Management Science* 59(1): 226-244.

Berné, C., et al. (2001). "The effect of variety-seeking on customer retention in services." *Journal of Retailing and Consumer Services* 8(6): 335-345.

Bolton, R. N. (1998). "A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction." *Marketing Science* 17(1): 45-65.

Boncheck, M. C., Sangeet, Paul (2013). "The age of social products." *Harvard Business Review* available at https://hbr.org/2013/10/the-age-of-social-products

Boudreau, K. J., et al. (2011). "Incentives and problem uncertainty in innovation contests: An empirical analysis." *Management Science* 57(5): 843-863.

Bowden, J. L.-H. (2009). "The process of customer engagement: A conceptual framework." *Journal of Marketing Theory and Practice* 17(1): 63-74.

Brodie, R. J., et al. (2011). "Customer engagement: conceptual domain, fundamental propositions, and implications for research." *Journal of Service Research*: 1094670511411703.

Chapman, P. M. (1997). Models of engagement: Intrinsically motivated interaction with multimedia learning software, University of Waterloo.

Chen, P.-Y. and L. M. Hitt (2002). "Measuring switching costs and the determinants of customer retention in Internet-enabled businesses: A study of the online brokerage industry." *Information Systems Research* 13(3): 255-274.

Cleves, M. (2008). An introduction to survival analysis using Stata, Stata Press.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, N.J., L. Erlbaum Associates.

Doan, A., et al. (2011). "Crowdsourcing systems on the world-wide web." *Communications of the ACM* 54(4): 86-96.

Ghose, A. and S. P. Han (2014). "Estimating demand for mobile applications in the new economy." *Management Science* 60(6): 1470-1488.

Gruen, T. W., et al. (2000). "Relationship marketing activities, commitment, and membership behaviors in professional associations." *Journal of Marketing* 64(3): 34-49.

Hackman, J. R. and G. R. Oldham (1976). "Motivation through the design of work: Test of a theory." *Organizational Behavior and Human Performance* 16(2): 250-279.

Howe, J. (2006). "The rise of crowdsourcing." *Wired Magazine* 14(6): 1-4.

Huang, Y., et al. (2012). "Crowdsourcing contests: A dynamic structural model of the impact of incentive structure on solution quality." working paper available at *http://aisel.aisnet.org/icis2012/proceedings/EconomicsValue/6/*

Huang, Y., et al. (2014). "Crowdsourcing new product ideas under consumer learning." *Management Science* 60(9): 2138-2159.

Isaac, M. S., et al. (2015). "Capturing consumer experiences: A context-specific Approach to measuring engagement." *Journal of Advertising Research*, Forthcoming.

Iyer, G. and Z. Katona (2015). "Competing for attention in social communication markets." *Management Science*. 62.8 (2015): 2304-2320

Jeppesen, L. B. and K. R. Lakhani (2010). "Marginality and problem-solving effectiveness in broadcast search." *Organization Science* 21(5): 1016-1033.

Kazai, G., et al. (2011). Crowdsourcing for book search evaluation: Impact of hit design on comparative system ranking. *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM.

Kim, Y. H., et al. (2013). "A Study of mobile user engagement (MoEN): Engagement motivations, perceived value, satisfaction, and continued engagement intention." *Decision Support Systems* 56: 361-370.

Kittur, A. and R. E. Kraut (2008). Harnessing the wisdom of crowds in Wikipedia: quality through coordination. *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ACM.

Klem, A. M. and J. P. Connell (2004). "Relationships matter: Linking teacher support to student engagement and achievement." *Journal of School Health* 74(7): 262-273.

Lee, D., et al. (2015). "Advertising content and consumer engagement on social media: evidence from Facebook." Available at SSRN 2290802.

Levine, S. S. and M. J. Prietula (2013). "Open collaboration for innovation: principles and performance." *Organization Science*.

Lewis, M. (2004). "The influence of loyalty programs and short-term promotions on customer retention." *Journal of Marketing Research* 41(3): 281-292.

Liu, C. Z., et al. (2012). "An empirical study of the freemium strategy for mobile apps: Evidence from the google play market." working paper available at *http://aisel.aisnet.org/icis2012/proceedings/EconomicsValue/14/*

McAlexander, J. H., et al. (2002). "Building brand community." *Journal of marketing* 66(1): 38-54.

McAlister, L. (1982). "A dynamic attribute satiation model of variety-seeking behavior." *Journal of Consumer research* 9(2): 141-150.

McAlister, L. and E. Pessemier (1982). "Variety seeking behavior: An interdisciplinary review." *Journal of Consumer Research* 9(3): 311-322.

Mollen, A. and H. Wilson (2010). "Engagement, telepresence and interactivity in online consumer experience: Reconciling scholastic and managerial perspectives." *Journal of Business Research* 63(9): 919-925.

Nishikawa, H., et al. (2013). "User-generated versus designer-generated products: A performance assessment at Muji." *International Journal of Research in Marketing* 30(2): 160-167.

Norton, M. I., et al. (2011). "The'IKEA effect': When labor leads to love." Harvard Business School Marketing Unit Working Paper(11-091) available at *http://www.hbs.edu/faculty/Publication%20Files/11-091.pdf*.

O'Brien, H. L. and E. G. Toms (2008). "What is user engagement? A conceptual framework for defining user engagement with technology." *Journal of the American Society for Information Science and Technology* 59(6): 938-955.

O'Brien, H. L. and E. G. Toms (2010). "The development and evaluation of a survey to measure user engagement." *Journal of the American Society for Information Science and Technology* 61(1): 50-69.

Oestreicher-Singer, G. and L. Zalmanson (2012). "Content or community? A digital business strategy for content providers in the social age." A Digital Business Strategy for Content Providers in the Social Age (July 01, 2012).

Pierce, J. L., et al. (2003). "The state of psychological ownership: Integrating and extending a century of research." *Review of general psychology* 7(1): 84.

Poetz, M. K. and M. Schreier (2012). "The value of crowdsourcing: can users really compete with professionals in generating new product ideas?" *Journal of Product Innovation Management* 29(2): 245-256.

Ransbotham, S. and G. C. Kane (2011). "Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia." *MIS Quarterly*-Management Information Systems 35(3): 613.

Ransbotham, S., et al. (2012). "Network characteristics and the value of collaborative user-generated content." *Marketing Science* 31(3): 387-405.

Ren, Y., et al. (2015). "The Impact and Evolution of Group Diversity in Online Open Collaboration." *Management Science* 62(6): 1668-1686.

Stewart, O., et al. (2009). Designing crowdsourcing community for the enterprise. Proceedings of the ACM SIGKDD *Workshop on Human Computation*, ACM.

Thomas, J. S. (2001). "A methodology for linking customer acquisition to customer retention." *Journal of Marketing Research* 38(2): 262-268.

Van Doorn, J., et al. (2010). "Customer engagement behavior: Theoretical foundations and research directions." *Journal of Service Research* 13(3): 253-266.

Van Trijp, H. C., et al. (1996). "Why switch? Product category: level explanations for true variety-seeking behavior." *Journal of Marketing Research*: 281-292.

Verhoef, P. C., et al. (2010). "Customer engagement as a new perspective in customer management." *Journal of Service Research* 13(3): 247-252.

Vivek, S. D. (2009). A scale of consumer engagement, The University of Alabama TUSCALOOSA.

Webster, J. and J. S. Ahuja (2006). "Enhancing the design of web navigation systems: The influence of user disorientation on engagement and performance." *MIS Quarterly* 30(3): 661-678.

Wooten, J. O. and K. T. Ulrich (2011). "Idea generation and the role of feedback: Evidence from field experiments with innovation tournaments." available at SSRN 1838733.

Wooten, J. O. and K. T. Ulrich (2013). "The Impact of Visibility in Innovation Tournaments: Evidence from Field Experiments." available at SSRN 2214952.

Zhang, Y., et al. (2016). "Modeling User Engagement in Mobile Content Consumption with

Tapstream Data and Field Experiment." available at SSRN 2847634.

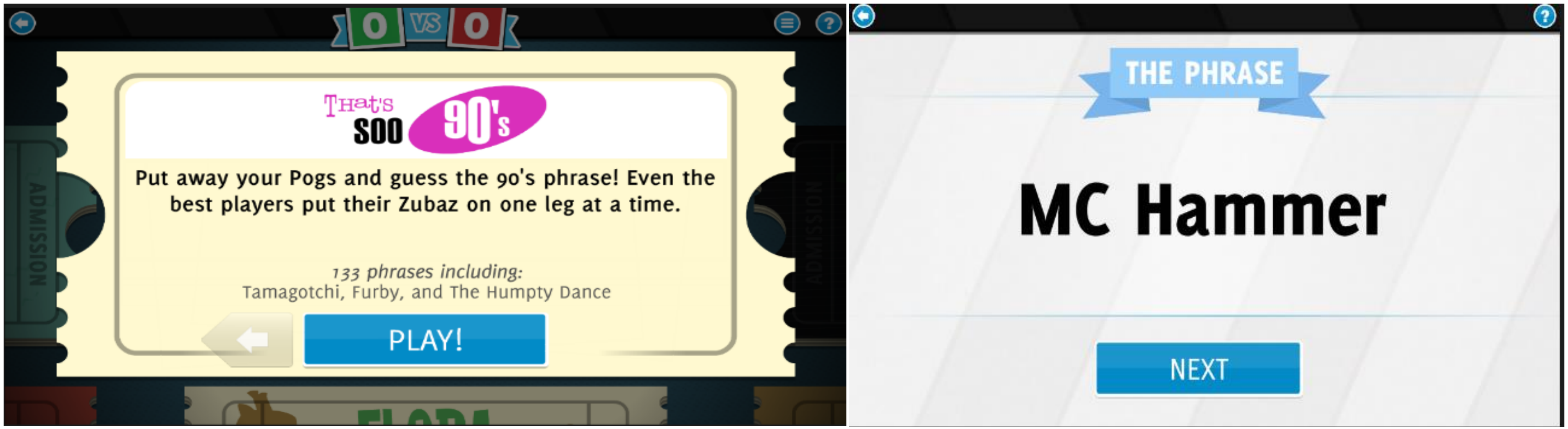
Appendix A. Brief of Experiment Design

Appendix A1. Game User Interface of Different Experiment Groups

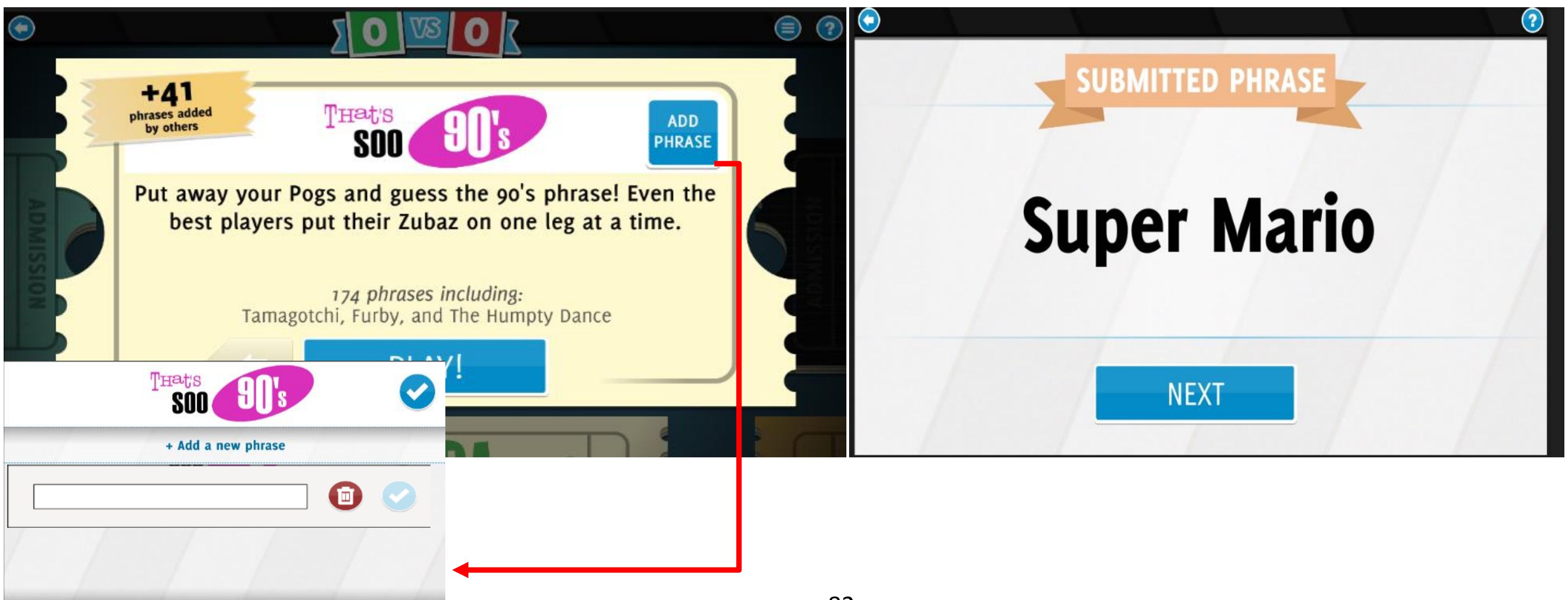
Figure A1.1 Comparison of Mobile Game User Interface between Different Experiment Groups

Card Deck Screen **Card Screen**

Control Group Version



Full Crowdsourcing Version



A2. Other Game Details

The application keeps track of all users' tapstream records from installation, word submission to game play. Tapstream data structure of a game session for a particular user is shown in **Table 2**. Each time a user opens the game app, the app records the beginning of a game session. The user will then be exposed to card decks of different topics. Once the user chooses a particular topic, she begins a game round. Words contained in the card deck flash one by one. Each time the user sees a word and click "Next" to confirm moving onto the next word phrase, a "SawWord" event will be recorded. Round duration assigned by game server follows a random distribution around a mean value of 60 seconds as is shown in **Figure A2.1**. Users will be prompted to input number of players involved before the game begins.

Figure A2.1. Distribution of Round Durations

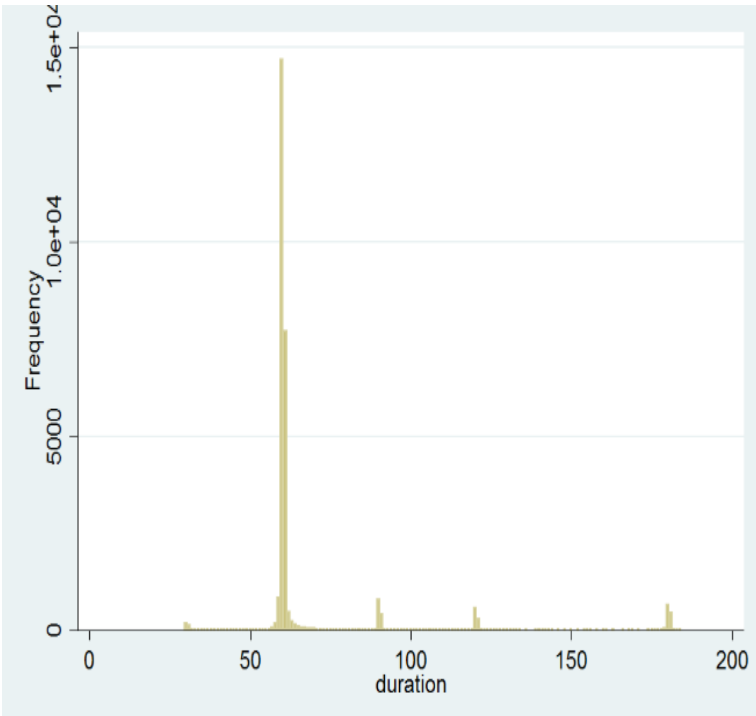


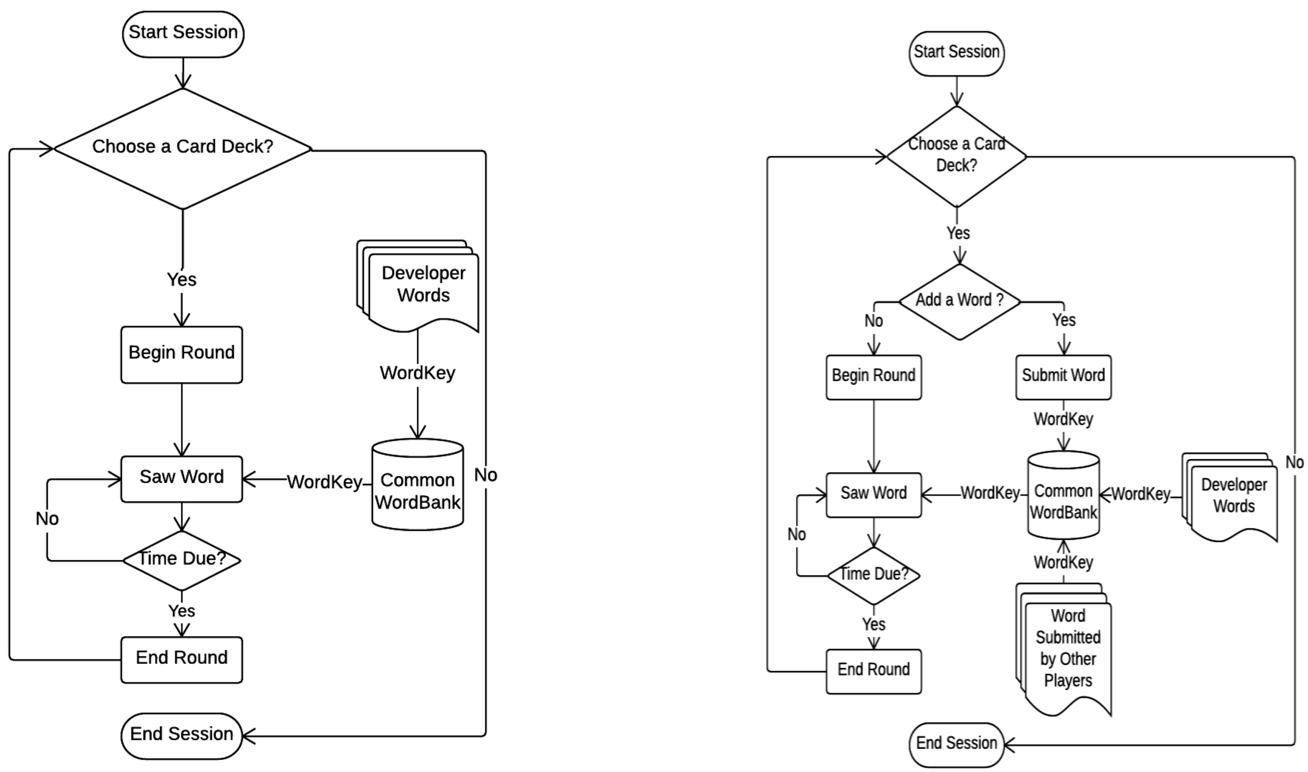
Table A2.1. Illustration of Game Events Hierarchy

Event	Attributes		TimeStamp
StartSession	Session1		2015-04-30
BeginRound	Round1		13:38:17.000
	Category: Spring Time		⋮
SawWord		sprout	⋮
SawWord		Easter bunny	
……		……	
SawWord		verdant	
EndRound			
……	……		
SubmitWord	Category: Music Terms		13:42:40.000
		Solo	
……	……		
BeginRound	Round2		13:53:40.000
	Category: Music Terms		⋮
SawWord		quartet	⋮
……		……	
SawWord		fermata	
EndRound			
EndSsion			

They have three options to choose from pop up window, namely, “2-4 players”, “5-7 players” and “8+ players”. According to our data, the average round duration assigned by the game server is longest when 5-7 players option is chosen and shortest when 2-4 players option is chosen.

Figure A2.2 Illustration of the Game Process of the Control and Treated Versions

Control **Full Crowdsourcing**



When a game round ends, users can choose to start the next round, or end the game session by exiting the game app. If the user forgets to exit the game app, the app will close the session after being left idle for 10 minutes. Sometimes users quit a game round before they finish the game, which is recorded as a “QuitRound” event. The rich details of the tapstream data allow us for a deep investigation into the mechanism through which crowdsourcing affects user behavior.

A3. Examine the Relationship between Game Round Duration and Number of Players

In **Table A3.1**, we pick durations of 30 seconds, 60 seconds, 90 seconds, 120 seconds, and 180 seconds, which are durations of peak frequency. We tabulate the frequency of these round durations over different number of players involved. The ANOVA test in **Table A3.2** and **Table A3.3** show that game rounds with more than 5 players have significantly longer duration than those with less than 5 players (In **Table A3.2**, **F = 19.81** rejects the assumption that groups with different number of players has the same mean value of round duration. In **Table A3.3**, games round groups with 5 to 7 players is on average 2.88 seconds longer than groups with 2 to 4 game players. And round groups with more than 8 players are 1.52 seconds longer than groups with 2 to 4 game players). However, game rounds with more than 8 players does not show a significant increase in duration compared to game rounds with 5 to 7 players.

Table A3.1. Frequency of Round Durations over Different Number of Players Involved
(select round duration = 30 seconds, 60 seconds, 90 seconds, 120 seconds, and 180 seconds)

	Duration					
numPlayer	**30s**	**60s**	**90s**	**120s**	**180s**	**Total**
2 to 4	135	11807	628	386	503	13459
5 to 7	29	1778	103	118	109	2137
8+	21	1105	56	72	45	1299
Total	185	14690	787	576	657	16895

Table A3.2. One-way ANOVA Analysis of Round Duration Variance
(Groups are defined by number of players involved in the game round.)

	Analysis of Variance				
Source	**SS**	**df**	**MS**	**F**	**Prob > F**
Between Groups	29635.59	2	14817.79	19.81	0.00
Within Groups	22108380.3	29550	748.17		
Total	22138015.8	29552	749.12		

Table A3.3. Comparison of Duration by Number of Players (Bonferroni)

Row Mean – Col Mean	**2 to 4**	**5 to 7**
5 to 7	2.88 (0.00)	
8+	1.52 (0.04)	-1.36 (0.19)

A4. Supplement to Pretreat Balance Test

Except for examining whether users are randomly assigned to the control and treated groups with regard to their installation demographics among, we also check whether new arrival users are assigned to different groups randomly overtime. In **Table A4.1,** we list each group's monthly installation. The percentage we listed represent how much percentage that month's installation account for the group's total installation during the whole experiment period. From the table, we can see that none of the groups have significantly different number user installation from other groups for each month.

We further run the one way Anova analysis to examine if there exists any mean or variance difference in users' daily installation across the treatment groups. From the Anova analysis results listed in Table A4.2, we can see that none of the F statistics or the Bartlett's Chi square statistics are significant, indicating no mean or variance difference for user daily installation across the groups over the months.

Table A4.1. User Installation over Time and Summarization of Rounds with Different Number of Players

		C	**F**	**S**	**A**	**F - C**	**S - C**	**A – C**
Installation Date		Count (percent)	Count (percent)	Count (percent)	Count (percent)	t-stat (SE)	t-stat (SE)	t-stat (SE)
Month	Day							
Feb	17 ~ 29	87 (11%)	94 (12%)	80 (11%)	94 (12%)	0.17 (0.02)	-0.44 (0.02)	0.54 (0.02)
Mar	1 ~ 31	223 (30%)	261 (33%)	245 (33%)	240 (32%)	1.37 (0.02)	1.49 (0.02)	0.91 (0.02)
Apr	1 ~ 30	207 (27%)	197 (24%)	185 (25%)	181 (24%)	-1.21 (0.02)	-1.06 (0.02)	-1.56 (0.02)
May	1 ~ 31	182 (24%)	191 (24%)	175 (23%)	179 (24%)	-0.01 (0.02)	-0.21 (0.02)	-0.21 (0.02)
Jun	1 ~ 9	58 (8%)	55 (7%)	57 (8%)	65 (8%)	-0.59 (0.01)	0.01 (0.01)	0.64 (0.01)
Number of Players		Round Count	Round Count	Round Count	Round Count	t-stat (SE)	t-stat (SE)	t-stat (SE)
2-4 Players		8.38 (19.97)	9.84 (19.09)	11.20 (25.91)	9.61 (24.10)	1.47 (0.99)	2.36 ** (1.19)	1.09 (1.14)
5-7 Players		1.30 (6.57)	1.38 (7.28)	1.67 (9.33)	1.44 (8.18)	0.21 (0.35)	0.89 (0.42)	0.36 (0.38)
8 + Players		0.87 (5.00)	0.76 (4.92)	1.39 (9.10)	0.56 (4.29)	-0.46 (0.25)	1.36 (0.38)	-1.32 (0.24)
Total Obs.		757	798	742	759	1555	1499	1516

Note: C, F, S, A in the first column is the abbreviation for the control and treated group names.
C: Control Group
F: Full Crowdsourcing Group
S: Submission-only Group
A: Access-only Group

Table A4.2 Daily Installation Difference among Treatment Groups Over the Months

	Anova Test Results			
Month	F(n1, n2)	Prob > F	Chi 2 (n1)	Prob > Chi 2
Feb 17 ~ Feb 29	0.26	0.85	0.01	1.00
Mar 1 ~ Mar 31	0.67	0.57	7.30	0.06
Apr 1 ~ Apr 30	0.69	0.56	2.00	0.57
May 1 ~ May 31	0.18	0.91	1.66	0.65
Jun 1~ Jun 9	0.23	0.88	3.51	0.32

Note: (n1, n2) is the degree of freedom for the F statistics. n1 is determined by the number of groups included in the Anova analysis. Since we have 1 control and 3 treated groups, n1 equals to 3. n2 depends on the total number of installation records of each months and thus varies over the months.

A5. Manipulation Test

To make sure that users treated with word submission options are aware of the existence of this word submission feature and understand how to use it, we run a manipulation test after we collect the data. The test results in Table A5.1 show that word contribution from users with word submission option are statistically significantly greater than zero. Also, t-test results in Table A5.2 and Table A5.3 show that people in the full-crowdsourcing group and the submission-only group contribute significantly greater number of words than users in control group whose word submission option is suppressed.

Table A5.1. Regression Result of Word Contributed

Word Contributed	Coef.	Std. Err.	t	p>|t|
submission allowed	0.20	0.37	5.24	0.00
_con	5.97	0.26	0.00	1.00

Table A5.2. T-test of Word Contribution Difference between Full-Crowdsourcing and Control Group

Group	Obs	Mean	Std. Err	Std. Dev.
Full-crowdsourcing	798	0.21	0.06	1.65
Control	757	0	0	0
combined	1555	0.11	0.03	1.18
diff		0.21	0.06	

diff = mean (Full-crowdsourcing) – mean (Control)		t = 3.58
H_0 : diff = 0		degrees of freedom = 1553
Ha: diff < 0	Ha: diff ≠ 0	Ha: diff > 0
Pr (T < t) = 0.9998	Pr (|T|>|t|) = 0.0004	Pr (T > t) =0.0002

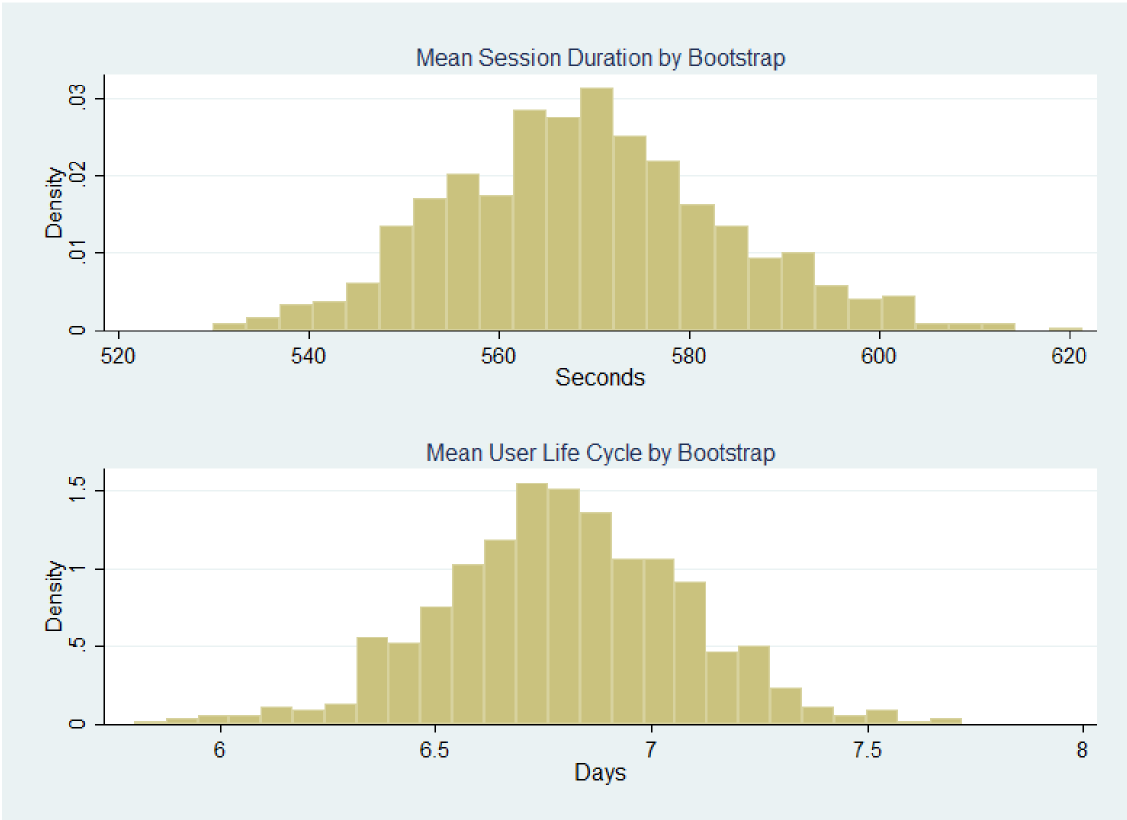
Table A5.3. T-test of Word Contribution Difference between Submission-only and Control Group

Group	Obs	Mean	Std. Err	Std. Dev.
Control	757	0	0	0
Submission-only	742	0.18	0.04	1.21
combined	1499	0.09	0.02	0.86
diff		0.18	0.04	

diff = mean (Submission-only) – mean (Control)		t = 3.97
H_0 : diff = 0		degrees of freedom = 1554
Ha: diff < 0	Ha: diff ≠ 0	Ha: diff > 0
Pr (T < t) = 1.0000	Pr (|T|>|t|) = 0.0001	Pr (T > t) = 0.0000

A6. Simulated Distribution of Sample Mean by Bootstrap

Figure A6.1 Histogram of Key Metrics Sample Mean Simulated by Bootstrap



Appendix B Supplement to Estimation Results

B1. Heterogeneous Effects

Table B1.1 Cox Hazard Model Estimation Results on Different User Sections

	Ending a Session		Abandoning the App	
	Top 25% User by TotalPlay	Lower 75% User by TotalPlay	Top 25% User by TotalPlay	Lower 75% User by TotalPlay
	1 Hazard ratio (p-value)	2 Hazard ratio (p-value)	3 Hazard ratio (p-value)	4 Hazard ratio (p-value)
Treatment				
Full-crowdsourcing	1.080 (0.287)	0.830*** (0.001)	0.633*** (0.001)	0.871* (0.052)
Submission-only	0.864** (0.044)	0.841** (0.002)	0.576*** (0.000)	0.947 (0.450)
Access-only	1.021 (0.775)	0.933 (0.213)	0.683** (0.005)	0.928 (0.299)
App Source				
Apple App Store	0.960 (0.599)	0.960 (0.587)	0.585** (0.003)	0.871 (0.161)
Google Play	1.019 (0.729)	1.022 (0.583)	0.949 (0.611)	0 .990 (0.850)
Geographic Location				
WestUS	1.028 (0.810)	0.881* (0.095)	0.913 (0.693)	0.850* (0.097)
MidWest-Central	0.977 (0.822)	0.928 (0.270)	0.958 (0.840)	0.958 (0.622)
EastUS	0.950 (0.632)	0.952 (0.473)	0.843 (0.431)	0.849* (0.067)
Number of Players				
5-7 Players	0.893 (0.133)	.959 (0.598)		
8+ Players	0.967 (0.727)	0.976 (0.770)		
Weekend	0.779*** (0.000)	0.968 (0.402)		
User Tenure Age (since installation)	1.007*** (0.000)	1.003* (0.059)		
Log Likelihood	-10908.046	-19694.296	-2464.3219	-10791.955
χ^2**(df)**	72.24*** (12)	23.47** (12)	30.39*** (8)	13.17* (8)
Obs	30432	6480	1695	2745
Subjects	1694	2745	747	2309

Notes: TotalPlay refers to user's total time spent on playing the game during their observed lifetime, which we use as a measurement for segment heavy players and casual players. A user is segmented into the top 25 percent or the lower 75 percent section according to the ranked Totalplay within her own treatment group. Users across different treatment groups are not lumped together and sorted for generating the top and lower segmentation. Since users are compared across treatment groups, there is no strong confounding issue between total play and our dependent variables here.

B2. Robustness Check

Table B2.1. Summary Statistics and Mean Comparison of TotalPlay under Different Treatments

Summary Stats of Player Metrics	**Control** (N = 757)	**Full-crowdsourcing** (N = 798)	**Submission-only** (N = 742)	**Access-only** (N = 759)
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
TotalPlay (in Seconds)	703 (1518.04)	814.366 (1589.884)	1006.55 (2291.902)	784.521 (1897.048)
Summary Stats of Player Metrics		(F – C) t-stats (SE)	(S – C) t-stats (SE)	(A – C) t-stats (SE)
Log (TotalPlay) (in Seconds)		2.55** (0.071)	2.39** (0.075)	1.00 (0.071)

Notes: The last three rows are the results of t-test calculating the mean difference between the control and the treated groups. "F – C" represents the difference between Full-crowdsourcing group and the Control group. "S – C", represents the difference between Submission-only group and the Control group. "A – C" represents the difference between Access-only group and Control group. TotalPlay is defined as the total time spent on the game by a user during his observed lifetime. Due to the highly skewed nature of the time span variables (e.g., session duration, user lifetime), we take log of these variables for t-test. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.001$

Table B2.2. Cox Proportional Hazard of Abandoning the Game App by Different Treatment Conditions using Different Censoring Period

Censoring Period	**=30 Days**		**= 38 Days**		**= 46 Days**	
	Treated Groups vs. Control	Treated Groups vs. Control (Control variables included)	Treated Groups vs. Control	Treated Groups vs. Control (Control variables included)	Treated Groups vs. Control	Treated Groups vs. Control (Control variables included)
	1 Hazard ratio (p-value)	2 Hazard ratio (p-value)	3 Hazard ratio (p-value)	4 Hazard ratio (p-value)	5 Hazard ratio (p-value)	6 Hazard ratio (p-value)
Treatment						
Full-crowdsourcing	0.838** (0.004)	0.836** (0.003)	0.834** (0.004)	0.832** (0.004)	0.843** (0.011)	0.842** (0.011)
Submission-only	0.855** (0.012)	0.853** (0.011)	.869** (0.030)	0.867** (0.028)	0.865** (0.035)	0.866** (0.036)
Access-only	0.877** (0.035)	0.868** (0.023)	.895* (0.086)	0.886* (0.061)	0.912 (0.175)	0.904 (0.137)
App Source						
Apple App Store		0.710*** (0.000)		0.676*** (0.000)		0.644*** (0.000)
Google Play		0.926* (0.099)		0.925* (0.109)		0.907* (0.056)
Geographic Location						
WestUS		0.822** (0.025)		0.810** (0.021)		0.809** (0.030)
MidWest-Central		0.860* (0.052)		0.869* (0.082)		0.919 (0.327)
EastUS		0.782** (0.002)		0.781** (0.003)		0.792** (0.009)
Log Likelihood	-15043.775	-15029.105	-13864.135	-13847.823	-12552.645	-12533.903
χ^2**(df)**	9.84** (3)	39.18*** (8)	8.72** (3)	41.34*** (8)	7.32* (3)	44.81*** (8)
Obs	4440	4440	4440	4440	4440	4440
Subjects	3056	3056	3056	3056	3056	3056

Notes: With censoring period equaling 38 Days, users whose last active session is less than 38 days away from the end of our experiment are considered to be censored. With censoring period equaling 46 Days, users whose last active session is less than 46 days away from the end of our experiment are considered to be censored. We choose these two censoring periods as more than 92% of the session intervals are within 38 days, and 95% of the session intervals are within 46 days. Users whose last observed active session is more than 38 or 46 days away from the end of our follow up period is not likely to come back to the app and is thus considered to have abandoned the app.

Table B2.3. Cox Hazard of Ending a Session Including Interaction Effects of Treatments with Number of Players, or Geographic Locations as a Robustness Check for Contamination

		Treatments Interacting with Number of Players		Treatments Interacting with Location WestUS	Treatments Interacting with Location MidWest-Central	Treatments Interacting with Location EastUS
		1 Hazard ratio (p-value)		2 Hazard ratio (p-value)	3 Hazard ratio (p-value)	4 Hazard ratio (p-value)
Treatment						
Full-crowdsourcing		0.967		0.985	0.915	0.979
		(0.472)		(0.751)	(0.114)	(0.691)
Submission-only		0.902**		0.908**	0.871**	0.871**
		(0.030)		(0.044)	(0.015)	(0.010)
Access-only		0.938		0.964	0.950	0.956
		(0.179)		(0.452)	(0.365)	(0405)
App Source						
Apple App Store		0.786***		0.791***	0.793***	0.794***
		(0.000)		(0.000)	(0.000)	(0.000)
Google Play		0.942*		0.943*	0.943*	0.943
		(0.061)		(0.069)	(0.066)	(0.069)
Geographic Location						
WestUS		0.900*		0.981	0.901*	0.900
		(0.095)		(0.847)	(0.095)	(0.092)
MidWest-Central		0.844**		0.844**	0.803**	0.842**
		(0.003)		(0.003)	(0.006)	(0.002)
EastUS		0.852**		0.851**	0.850**	0.848**
		(0.005)		(0.005)	(0.005)	(0.047)
Number of Players						
5-7 Players		0.891		0.844**	0.847**	0.845
		(0.297)		(0.002)	(0.002)	(0.002)
8+ Players		0.893		0.943	0.945	0.943
		(0.366)		(0.344)	(0.368)	(0.348)
Weekend		0.833***		0.836***	0.836***	0.835***
		(0.000)		(0.000)	(0.000)	(0.000)
User Tenure Age		0.997**		0.997**	0.997**	0.997**
(since installation)		(0.013)		(0.009)	(0.008)	(0.010)
Interaction Effects						
<u>Full</u>	*5-7 Players	0.890	* WestUS	0.885		
		(0.450)		(0.289)		
	*8+ Players	1.135	* Mid_Central		1.138	
		(0.472)			(0.137)	
			* EastUS			0.958
						(0.643)
<u>Submission</u>	*5-7 Players	0.867	* WestUS	0.871		
		(0.353)		(0.237)		
	*8+ Players	0.950	* Mid_Central		1.045	
		(0.765)			(0.621)	
			* EastUS			1.059
						(0.533)
<u>Access</u>	*5-7 Players	1.065	* WestUS	0.934		
		(0.680)		(0.556)		
	*8+ Players	1.204	* Mid_Central		1.009	
		(0.312)			(0.922)	
			* EastUS			0.993
						(0.943)
Log Likelihood		-33269.917		-33271.369	-33270.78	-33271.567
χ^2**(df)**		102.16 ***		99.26 ***	100.44 ***	98.87 ***
		(18)		(15)	(15)	(15)
Obs (# of Rounds)		36912		36912	36912	16797
Subjects (# of Sessions)		4439		4439	4439	4439

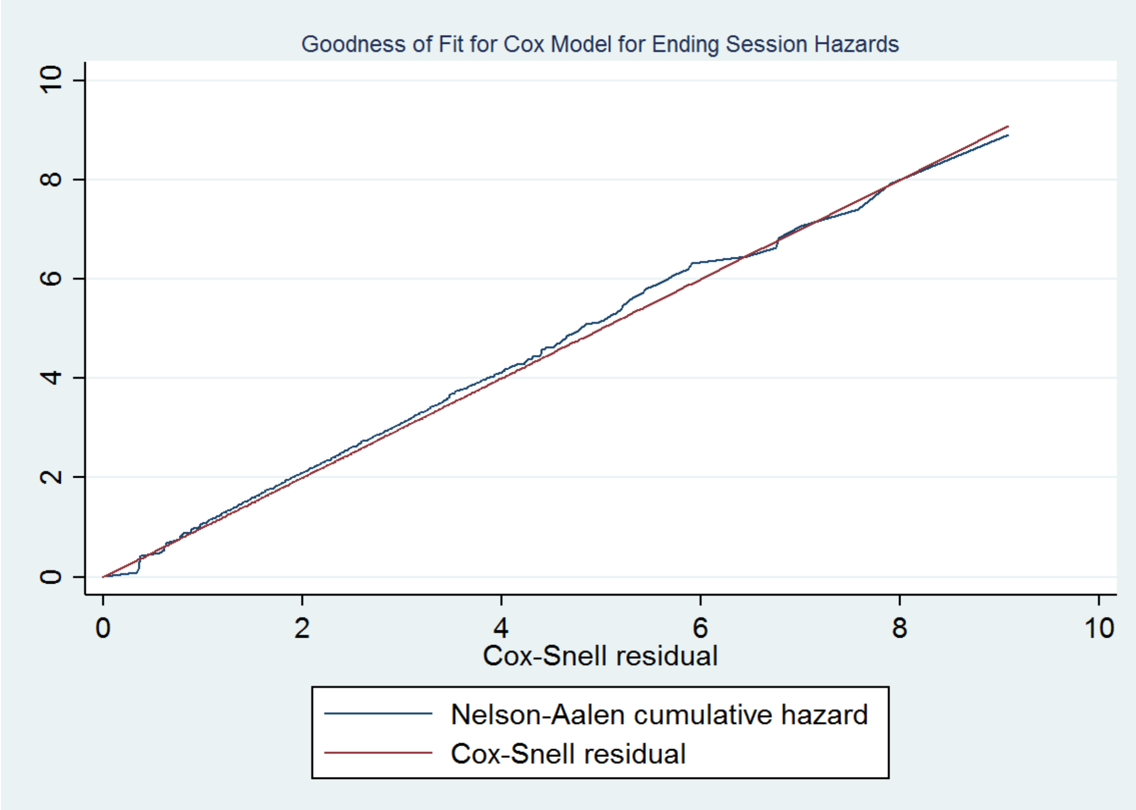
Table B2.4. Cox Hazard of Abandoning an App Including Interaction Effects of Treatments with Number of Players, or Geographic Locations as a Robustness Check for Contamination

		Treatments Interacting with Location WestUS	Treatments Interacting with Location MidWest-Central	Treatments Interacting with Location EastUS
		2 Hazard ratio (p-value)	3 Hazard ratio (p-value)	4 Hazard ratio (p-value)
Treatment				
Full-crowdsourcing		0.837** (0.009)	0.833** (0.028)	0.815** (0.006)
Submission-only		0.838** (0.012)	0.856* (0.064)	0.867* (0.059)
Access-only		0.902 (0.139)	0.856* (0.065)	0.845 ** (0.027)
App Source				
Apple App Store		0.680*** (0.000)	0.679*** (0.000)	0.678*** (0.000)
Google Play		0.925* (0.100)	0.925* (0.100)	0.927 (0.107)
Geographic Location				
WestUS		0.842 (0.211)	0.810** (0.019)	0.812** (0.020)
MidWest-Central		0.858* (0.053)	0.854 (0.153)	0.860** (0.057)
EastUS		0.777** (0.002)	0.778** (0.002)	0.812** (0.020)
Interaction Effects				
Full	* WestUS	0.942 (0.722)		
	* Mid_Central		0.988 (0.924)	
	* EastUS			1.056 (0.689)
Submission	* WestUS	1.133 (0.983)		
	* Mid_Central		1.003 (0.621)	
	* EastUS			0.966 (0.805)
Access	* WestUS	0.802 (0.197)		
	* Mid_Central		1.035 (0.787)	
	* EastUS			1.094 (0.516)
Log Likelihood		-14351.222	-14353.331	-14352.926
χ^2**(df)**		47.19 *** (11)	42.97 *** (11)	43.79 *** (11)
Obs (# of Sessions)		4440	4440	4440
Subjects (# of Users)		3056	3056	3056

Table B2.5. Cox Hazard of Abandoning an App Including Interaction Effects of Treatments with Number of Players, or Geographic Locations as a Robustness Check for Contamination

		Treatments Interacting with Location WestUS	Treatments Interacting with Location MidWest-Central	Treatments Interacting with Location EastUS
		2 Hazard ratio (p-value)	3 Hazard ratio (p-value)	4 Hazard ratio (p-value)
Treatment				
Full-crowdsourcing		0.837** (0.009)	0.833** (0.028)	0.815** (0.006)
Submission-only		0.838** (0.012)	0.856* (0.064)	0.867* (0.059)
Access-only		0.902 (0.139)	0.856* (0.065)	0.845 ** (0.027)
App Source				
Apple App Store		0.680*** (0.000)	0.679*** (0.000)	0.678*** (0.000)
Google Play		0.925* (0.100)	0.925* (0.100)	0.927 (0.107)
Geographic Location				
WestUS		0.842 (0.211)	0.810** (0.019)	0.812** (0.020)
MidWest-Central		0.858* (0.053)	0.854 (0.153)	0.860** (0.057)
EastUS		0.777** (0.002)	0.778** (0.002)	0.812** (0.020)
Interaction Effects				
Full	* WestUS	0.942 (0.722)		
	* Mid_Central		0.988 (0.924)	
	* EastUS			1.056 (0.689)
Submission	* WestUS	1.133 (0.983)		
	* Mid_Central		1.003 (0.621)	
	* EastUS			0.966 (0.805)
Access	* WestUS	0.802 (0.197)		
	* Mid_Central		1.035 (0.787)	
	* EastUS			1.094 (0.516)
Log Likelihood		-14351.222	-14353.331	-14352.926
χ^2**(df)**		47.19 *** (11)	42.97 *** (11)	43.79 *** (11)
Obs (# of Sessions)		4440	4440	4440
Subjects (# of Users)		3056	3056	3056

Figure B2.1 Test the Overall Goodness of Fit for using Cox Hazard Model



B3. Discrete Choice Model for the Probability of SawCustomWord Event

Table B3.1. Estimated Probability of SawCustomerWord Event for the Full-crowdsourcing and Submission-only Group

Saw Custom Word	Odds Ratio	Std. Err.	P>|z|
Word Submission	616.447	368.855	0.000
Submission-only	2.096	.658	0.018
Round Duration	0.999	.003	0.683
_cons	.000	.000	0.000

Notes:

We include only the submission-only group and the full-crowdsourcing group in the logit regression.

"Saw Custom Word" = 1: the user saw custom words in the current game round; otherwise "Saw Custom Word" = 0.

Submission-only = 1: the user is in the submission-only group; otherwise, Submission-only = 0.

The regression result indicates that after controlling the word submission behavior and the round duration, a user in the submission-only group is two times higher than a user in the full-crowdsourcing group.