THE TWOFOLD EFFECT OF CUSTOMER RETENTION IN FREEMIUM SETTINGS

Eva Ascarza
Harvard Business School

Oded Netzer
Columbia Business School

Julian Runge
Humboldt University Berlin

November 2020

Eva Ascarza is the Jakurski Family Associate Professor of Business Administration, Harvard Business School (email: eascarza@hbs.edu). Oded Netzer is the Arthur J. Samberg Professor of Business at Columbia Business School (email: onetzer@gsb.columbia.edu), and Julian Runge is researcher at the Institute of Marketing of Humboldt University Berlin (email: julian.runge@hu-berlin.de). The authors are grateful to Hengyu Kuang for excellent research assistantship.

THE TWOFOLD EFFECT OF CUSTOMER RETENTION IN FREEMIUM SETTINGS

Abstract

The main tradeoff in designing freemium services is how much of the product to offer for free. At the heart of such a tradeoff is the balancing act of providing a valuable free product in order to acquire and engage consumers, while making the free product limited enough to entice some consumers to pay for the service. We argue that customer retention plays an important role in balancing this tradeoff, due to customer retention's twofold effect in freemium services—it not only generates additional revenues from the advertising side but can also increase user's monetization in the future—firms are often myopic in their retention efforts by not recognizing that by limiting the free product, the firm is not only risking monetary losses from the advertising side but is also suffering monetary losses from the premium side.

We leverage a large-scale field experiment run in a mobile game context, where game difficulty was exogenously manipulated among customers at risk of churning. As expected, giving customers an easier game significantly decreases purchases in the specific round played. However, lowering the game difficulty not only increases short-term play and subsequent retention, but also increases customer spending from premium services both in the short- and in the long-run. We find substantial heterogeneity in the effect of reducing game difficulty, where customers who are more prone to making progress in the games exhibit stronger effects and customers who previously spent money on the game exhibit the strongest effect on in-app purchases. We leverage these insights to recommend personalized freemium product design strategies and show how the focal firm can further increase revenues from both advertising and premium services.

Keywords: Freemium, Retention, Monetization, Gaming, Field Experiments

1 INTRODUCTION

Due to minimal marginal cost of serving additional customers and the opportunity to reach large numbers of people on digital platforms (Forbes 2015; Kumar 2014), "freemium" has become a widespread pricing model for the sale of digital goods. Many digital products such as media and information websites, music platforms, online games, dating apps and social network sites offer part of the product for free (often accompanied by advertising) and an enhanced customer experience for monetary payment. In freemium services, revenue primarily arises from two sources: money from the free part (via advertising), and money from the premium side via payments for extras.

The main tradeoff frequently discussed in the context of freemium is how much of the product to offer for free (Cheng and Tang 2010; Halbheer et al. 2014). At the heart of the tradeoff is that firms may wish to provide a valuable free product in order to engage consumers in the free version of the product to generate revenue through advertising and to entice new customers to try out the product, but at the same time the firm wishes to limit the free product it offers to encourage consumers to pay to enhance their experience. For example, in the context of free-to-play (F2P) games, the premium service comes in the form of game "add-ons" (e.g., virtual tools, puzzle cues, extra time) that enhance the game, enabling the player to make progress in the game, or to win against other players. Game designers can then balance the tradeoff between the free and paid premium aspects of the game by, for example, adjusting the game's difficulty. Reducing game difficulty can be thought of as increasing the free portion of the product, as consumers can buy extras to make the game easier in order to pass a certain level.

This tradeoff between free and premium has traditionally been discussed in the context of customer acquisition and short-term revenue, with a great focus on the network externality effect offered by the adopters of the service (e.g., Shi, Zang and Srinivasan 2019). However, little attention has been given to the implications of this tradeoff in users' subsequent engagement and retention. On the one hand, the firm can increase customer retention (prevent customer churn) by offering to customers a better experience for free—and therefore increase the number of active customers in the long-term which is associated with higher advertising revenue. However, by providing more of the free product, the firm may cannibalize customers' spending on premium services. Striking the balance in this tradeoff can be challenging. First, it is difficult for established firms to exogenously manipulate customer retention without risking possible losses in profitability (from customers who might react negatively to the change). As such, firms are generally reluctant to take actions that can alter customers' retention dramatically. Second, preventing churn requires long-term counterfactual thinking. That is, the firm needs to estimate what would have been the churn without the intervention, and often wait for long periods of time to properly measure the long-term consequences.[1]

The purpose of this research is to bring in retention and retention prevention into the discussion of the balance between free and premium in the freemium economy. We investigate the consequences of adjusting the free vs. premium tradeoff on customer behavior and argue that, from a retention prevention perspective, firms are often myopic by not recognizing that by limiting the free product, the firm is not only risking monetary loses from the advertising side but also may suffer monetary loses from the premium side as the reduction of the "free" portion of the product may increase customer churn, and these churned customers will never have the chance to buy premium products. We argue that, because customer retention can have a twofold

[1] New firms (or startups) are generally prone to experiment with the product or price and see the impact in retention. However, those firms usually lack the size (their customer base is usually small) and the persistence (their product keep changing constantly) to measure long-term consequences of the actions that affected retention in the first place.

effect in freemium services—it not only generates additional revenues from the advertising side but can also increase user's monetization in the future—freemium services can turn this difficult tradeoff into a win-win situation for customers at the risk of churning where adjustments in the balance between the free/premium design for customers at risk of churn can increase customer retention *and* increase profitability as well. This is of great importance for freemium contexts and particularly gaming where churn rates are unusually high and a large proportion (in the case of this game nearly half of the customers) are at risk of churn.

Managing churn is especially challenging in freemium services because in these settings firms do not have many levers in order to persuade users to stay. In domains outside of the freemium economy, the main lever companies use in proactive or reactive retention campaigns is to offer customers discounts (Ascarza et al. 2018). However, given that most freemium customers enjoy the product for free, discounts rarely alter users' behavior (Levitt et al. 2016) and therefore firms in these contexts need to use other levers. As we illustrate in this research, changing the product itself by adjusting the free/premium tradeoff is such a retention lever.

We show the twofold effect of retention in a freemium game context whereby difficulty of the game—and therefore the player's ability to make progress within the game—can be manipulated. Specifically, the difficulty of the game can be increased in order to encourage users to make in-app purchases (which will help them to pass difficult rounds) or can be decreased in order to increase engagement and retention (as users often enjoy making progress in the game). We track the behavior of over 300,000 players over the course of 12 weeks through which game difficulty was exogenously manipulated through a field experiment. When users were identified as being at a high risk of churning (proxied by infrequent playing), they were assigned to treatment—in which case their game difficulty was reduced—or to control—whose game difficulty remained the same. This exogenous variation in the level of difficulty allows us to assess the effect of the ability to make progress on customer behavior. Specifically, we measure the short- and long-term effect of the intervention on customer engagement, retention, and purchases. We find that, while giving customers an easier game significantly decreases purchases in the specific game round played—this effect is to be expected because treated users are getting more progress for "free"—*lowering the game difficulty* not only increases (short-term) play and subsequent retention increasing revenue from advertising, but also *increases customer spending* from premium services in subsequent rounds. The increase of premium purchases comes from two main sources: 1) in the long-term, treated customers (with lower game difficulty) buy more simply because they are more likely to be active, and 2) in the short-term, because treated customers are more engaged, they tend to buy extras that can expedite wait time in the game.

Using both heterogeneity in treatment effect and instrumental variable regressions of game progress on customer behavior (using treatment as an exogenous instrument), we show that the mechanism for the positive effect of reduced game difficulty on engagement, retention, and monetization, is through consumers who strive for game progress. Indeed, we find substantial heterogeneity in the effect of reducing game difficulty, where customers who are more prone to making progress in the games, and at times of frustration or when players can make more progress (farther away from gates) exhibit a stronger positive effect of reducing game difficulty. Additionally, customers who previously spent money on the game exhibit the strongest effect on in-app purchases. Our results underscore the importance of evaluating freemium products from a customer relationship perspective and that managing customer retention in the context of the freemium economy is important yet complex.

Our paper makes three main contributions. First, much of the research on freemium products and services focuses on the short-term tradeoff between customer acquisition or advertising revenue from the free product and the payment for the premium service (e.g., Halbheer et al. 2014, Appel et al. 2020, Lambrecht and Misra 2017). In this research, we focus both on the short- and on the long-term effects of this tradeoff, paying special attention to the effect of the design of the freemium product on customer relationship management and retention efforts. Second, this is one of the first studies to empirically and exogenously manipulate the user experience that is at the core of the free vs. fee tradeoff—in the context of the game, the users' ability to make progress under the free version of the service. We explore the use of game difficulty as a lever for customer engagement, retention and profitability in mobile games. Mobile gaming is one of the fastest growing industries and therefore understanding consumer retention in this context is of growing importance for marketers. Our experimental setup allows us to measure the causal effect of changing game difficulty on game retention and (short- and long-term) profitability. Finally, from a substantive point of view, we provide actionable advice to managers in freemium settings such as mobile apps and gaming. Specifically, we claim that, companies in freemium services may be leaving money on the table by not taking into account the design of the product itself (the free-premium balance) and the twofold effect of retention in the form of increased advertising as well as increased likelihood of future purchases in their retention efforts.

2 PREVIOUS LITERATURE

This paper is related to several streams of literature including freemium services, customer retention, gaming, and the growing body of work that investigates the impact of firms' interventions on customer-side outcomes.

First, our study is closely related to the growing literature on freemium services. One can split freemium products and services into two types. One in which customers get free access to a reduced version of the product, but can pay a subscription or a one-time fee to "unlock" a full, enhanced, or improved (e.g., advertising free) experience of the product, often called feature-limit freemium (e.g., Dropbox or newspapers). The second type of freemium, mainly employed in mobile apps and games (Liu et al. 2014), offers a free version of the product and allows users to buy items throughout the use of the product to improve the user experience often called the in-app-purchases (IAP) model (e.g., buy avatar or buy coins to eliminate required wait time in game progress). In this paper we focus on the latter type (IAP) which is employed in 79% of the mobile games (and 50% of the mobile non-game apps), and accounts for almost half of the total mobile gaming market.[2]

A set of analytical and empirical studies have investigated the economic viability of the freemium model relative to its two extreme cases: free only or paid only (Chiou and Tucker 2013; Halbheer et al. 2014; Li, Jain and Kannan 2019; Liu et al. 2014; Shi, Zhang and Srinivasan 2019). A major driver behind the economic viability of adding a free version to the paid product is consumers' reduced cost to sample the product and reduced uncertainty about the product's quality (Appel et al. 2020; Deng, Lambrecht and Liu 2020; Li, Jain and Kannan 2019). Several papers have focused on quantifying the amount of free product that should be offered. Lambrecht and Misra (2017) suggest that the amount of free product should depend on the demand for the product, proposing a "countercyclical offering." Gu, Kannan and Ma (2018) expand the view of the product design between free and premium by investigating the impact of line extensions in

[2] https://www.statista.com/statistics/273120/share-of-worldwide-mobile-app-revenues-by-channel/
https://www.statista.com/statistics/297024/most-popular-mobile-app-monetization-models/
https://techjury.net/stats-about/mobile-gaming/

the context of freemium. They show how extending the product line of the premium product could increase demand for the existing premium version, thus increasing overall revenues.

A number of papers from information systems have focused on developing models to identify customers who are most likely to convert for free to premium (e.g., Sifa et al. 2018; Voigt and Hinz 2016). An important feature of conversion had been attributed to social effects (Oestreicher-Singer and Zalmanson 2013; Shi, Zang and Srinivasan 2019), peer influence (Bapna and Umyarov 2015) and referrals (Lee, Kumar and Gupta 2017). Whereas network externality (e.g., Shi, Zang and Srinivasan 2019) have often been suggested as a reason to offer a free version of the product to expedite product adoption, such network forces are less relevant in freemium services such as our focal game, where the product/service has minimal social components.[3]

Most papers mentioned above have looked at feature-limited formats of freemium, in which customers consume a free and degraded version of the product, and a fraction of these customers transition to a paid version of the product. These papers tend to focus on the drivers of customers conversion from free to premium (e.g., Lee, Kumar and Gupta 2017). In this work, we investigate the IAP version of freemium in which the customer can use the app (play the game) for free but can enhance their experience by buying extras at any point in time. In such a format, because the use of premium is a "per transaction" behavior (rather than a switch to a paid subscription) other aspects of customer lifetime value such as recurrent usage, monetary purchases, and retention are key to understanding the profitability of the freemium design.

[3] Our game is a single-player game in which users cannot compete among each other. While users can, in theory, link their account with their Facebook profiles in order to see the scores of their friends, our data show that more than 90% of users never activate such a feature.

However, despite its potential importance the impact of freemium on customer retention has been largely understudied. In one exception, Appel et al. (2020) build an analytical model in which satiation (a parameter in their model that captures the customer's likelihood of being satiated from the content and therefore churn) can affect the design of the freemium product. In that paper, Appel and colleagues look at satiation as an exogenous aspect of the product, whereas we treat customer retention (or churn) as a direct outcome of the retention efforts and therefore assess the impact of the freemium product (dynamic) design on customer retention. Several computer science and information systems papers (e.g., Hadiji et al. 2014; Lee et al. 2016; Runge et al. 2014; Viljanen et al. 2020; Xie et al. 2015) use machine learning classifiers to predict churn in free to play games. In a context similar to ours, Runge et al. (2014) examine the effect of firm communications with at risk customers on decreasing churn of these customers, and find that, while the communication efforts are more successful under predictive churn management, it is hard to devise effective retention incentives at the end of customers' lifetime. The authors call for "A thorough assessment of the best strategy to deal with churning players in casual social games", and particularly propose "…to touch the deeper gameplay mechanics and change players' game experience in a way that keeps them interested in the game." (Runge et al. (2014), p. 8).

Thus, while previous work on freemium has recognized the tradeoff between free and premium offerings to maximize the short-term revenue from the free product (advertising) and paid product, to the best of our knowledge, no work has investigated the short- and long-term effect of user retention and retention efforts via the revenue generated by each of these components separately in the understudied context of IAP freemium.

Our work is also related to research that has used field experiments to investigate the causal link between the firm interventions and the user behavior (e.g., Ascarza 2018; Ascarza, Iyengar and Schleicher 2016; Godinho de Matos, Ferreira and Bello 2018; Lemmens and Gupta 2020; Levitt et al. 2016, Runge et al. 2019, Sun, T., Gao, G., & Jin, G. Z. 2019; Sun, Viswanathan and Zheleva 2020). Most literature in that domain has focused on interventions that relate to pricing, promotion, or communication decisions. Price and promotion incentives are difficult in freemium products (particularly at the free part of the product). Accordingly, our work focuses on product intervention by changing the attributes of the product itself (game difficulty) for different consumers.

In terms of domain of application, our work is also related to the work on video and mobile games and, in particular, to those related to customer engagement and lifetime play (e.g., Huang, Jasin, and Manchanda 2019). Amabile and Kramer (2011) build on their progress theory, to suggest that to maximize player engagement, game designers should make sure that players, 1) experience everyday progress, 2) experience small wins even in setbacks, and 3) were making progress in multiple ways. The ability to make progress and have small wins is at the heart of our experimental manipulation to reduce game difficulty. Indeed, Hofacker et al. (2016), suggested that striking the right balance of game difficulty and enhancing the relationship between game difficulty and reward could lead to positive game outcomes, and enhanced purchase intentions and unplanned purchases.

To summarize, our work contributes to the freemium economy literature, by investigating both the short- and long-term effect of the freemium product design on usage, IAPs and retention. We contribute to the more general work of firm interventions but going beyond the typical 2Ps of firm intervention (Price and Promotion) and towards intervention in the third P (Product design).

3 EMPIRICAL APPLICATION

In this section we present the empirical context, describing the nature of the business setting we investigate, the details about the freemium model, and present model-free evidence for the effect of game difficulty on game engagement, retention and monetization. We then describe the field experiment and the data we use in our empirical application.

3.1 Business context

We partnered with a company that develops mobile gaming applications. Mobile gaming is one of the fastest growing markets in today's digital market, accounting for more than 70% of consumer expenditure in mobile apps,[4] and expected to reach over $106 Billion in revenue worldwide in 2021.[5] This product category is of particular interest for our research because, the freemium model is particularly prevalent in the gaming industry with more than 75% of the game revenue coming from F2P games via IAP.[6] Furthermore, both user retention and monetization are key challenges in mobile applications (both gaming and non-gaming), providing us with an ideal context to investigate the tradeoffs and synergies of user retention and monetization in freemium contexts.

3.1.1 A free-to-play (F2P) game

The game we use for our empirical application is a F2P puzzle mobile game similar to the well-known Candy Crush Saga. In the game, players complete levels by combining three or more pieces of the same color laid out on a game board. Once the pieces are successfully combined, they disappear and the freed-up space on the game board is filled with new pieces. The more

[4] https://www.appannie.com/en/insights/market-data/games-accounted-for-70-of-consumer-spend-in-apps-in-q3-2019/
[5] https://mediakix.com/blog/mobile-gaming-industry-statistics-market-revenue/
[6] https://www.statista.com/statistics/413384/f2p-p2p-gaming-revenues-distribution-worldwide/

pieces the user manages to combine in each move, the more points they obtain. The game consists of many levels, which must be completed in sequence (i.e., a user cannot play level 33 until successfully passing level 32). Levels have various goals/challenges (e.g., combining a specific number of pieces of certain colors, or achieving a certain score) that must be completed within a fixed number of moves or limited amount of time. If the player does not meet the goal, the user loses one life and they must try again before continuing to the next level. When a user meets the level's goal, they receive one to three stars (based on the performance in that round), and are then allowed to play the next level.

The game has embedded mechanisms that systematically prevent customers from making progress. First, users have limited number of lives at any point in time. Specifically, players are given five lives (i.e., they have a "stock" of five lives). If a user loses all the lives, the user cannot continue playing until a new life is added to their stock, the game generates one life every 30 minutes, with a maximum of five lives in the stock at any point in time. Second, the game includes "a gate" every 20th level, starting at level 40. This means that, once a user successfully completes level 40, 60, 80, and so forth, there is a waiting period of five days, during which the player is not allowed to play higher levels (i.e., the player is not able to make any progress). At any point in the game, either being stopped at a gate or not, users can play lower levels, conditional on having lives.

3.1.2 Monetization via in-app-purchases (IAP)

The game monetizes users by selling game currency, aka "coins," which can be bought with real money. At the beginning of the game, users are given an endowment of 70 coins, which can be increased or replenished with real money. Users redeem coins to get one of three benefits. First, if a user "runs out of lives," the user can pay for a new life in order to continue playing. Second, when stopped at a gate, a player can use game currency to break the gate and continue playing higher levels.[7] Finally, users can use game currency to buy "extras" that help them achieve the goal of the level. For example, a user can buy a booster-touch that changes the color of a puzzle piece, allowing the player to combine more pieces, and therefore making more progress towards the level objectives and obtaining a higher score. Practically, these "extras" serve as a mechanism for users to pay in order to decrease the game difficulty for the played level. In essence, all paid items (extras, gates, lives) are instruments that enhance the user's ability to continue making progress in the game.

Similar to the vast majority of freemium products and mobile games in the market, most players never make a purchase. However, when they do, they tend to transact multiple times,[8] generating a sizeable amount of revenue for the focal firm.[9] In our game, most of the revenue from IAP comes from users buying extras and breaking gates, rather than paying for getting extra lives. To give a sense of price paid by the players, the average purchase amount (given purchase) is around $4 across all players, and the currency exchange to break a gate corresponds to $1.

3.1.3 Monetization via in-app advertising

The mobile game also makes money from advertising, which is directly related to the number of players, playing at any point time ("eyeballs"). In-app advertising is a fast growing market with predictions that would hit $240B in 2020.[10] The most common form of advertising consumption in this game are short (10-30 second) videos that users need to watch before they are allowed to

[7] Users can add one life to their "stock" or "break" a gate by sending requests to friends via Facebook. However, at the time of the data, this feature was rarely used—less than 10% of our uses even linked their game to Facebook—and therefore we ignore the activity on Facebook.
[8] Using as reference the focal game, holding the level of playing constant, the probability of making a purchase in a particular day is 17 times larger for users who have purchased before than for users who have never spent any money in the game.
[9] Actual monetization figures for the focal company are omitted to preserve confidentiality.
[10] https://www.appannie.com/en/go/state-of-mobile-2020/

play another round. This practice is very common among other mobile games, with 94% of game developers reporting to use in-app advertising in 2019.[11] While we do not have precise data on advertising exposure and revenues of our focal company (the company did not share such information), ceteris paribus, the more rounds the user plays, the higher the exposure to advertising and therefore the higher the advertising revenue generated.[12] As a result, player engagement is a key metric for the firm, one that is regularly monitored and improved upon (Seufert 2013).

3.2 User behavior

3.2.1 User retention, engagement, and monetization

Like most other freemium services such as F2P games and mobile apps, retention rates are rather low. Figure 1(left) shows, for a random sample of users (N=10,000) of this game, the proportion of users who play the game X days after installation (up to 28 days). As can be seen from the figure, only 47.54% of users play the next day after installation, and only 9.02% do so after 28 days.[13] Among those who play, the level of engagement with the game also declines over time. Figure 1(right) shows the average number of rounds played, conditional on the user playing at all on a particular day. This steep decline in retention and engagement highlights the importance for the firm to retain customers and keep them engaged.

[11] https://www.mobilemarketer.com/news/study-94-of-free-mobile-games-have-in-game-ads-as-developers-fortify-stra/569500/

[12] For example, a recent study found that ad impressions in the US mobile gaming market surged 57% over the pandemic period, resulting in the ad revenue for April being 59% greater than a year earlier (https://www.mobilemarketer.com/news/mobile-game-ad-revenue-jumps-59-during-pandemic/579516/).

[13] While some of these users can still play in future days, it is rarely the case that customers come back after several weeks of inactivity (Runge et al. 2014).

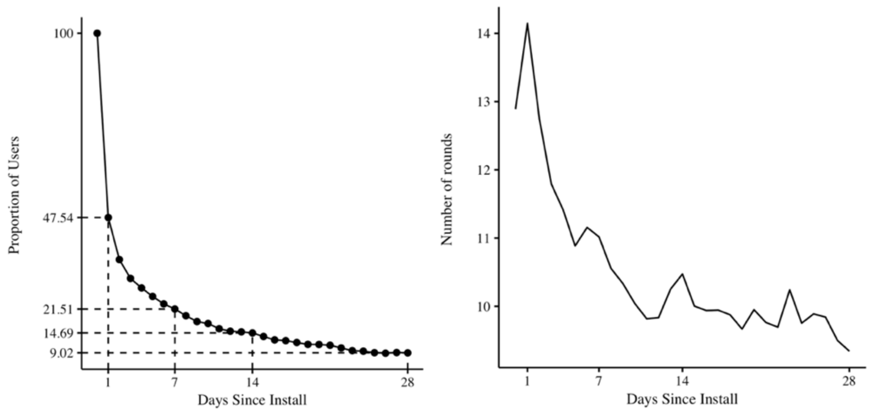


***Figure 1: Retention and Engagement during the days after installation**. The figure on the left shows the proportion of users who play up to 28 days after having installed the game. The figure on the right shows the average number of rounds played up to 28 days after installation, conditional on the user playing on that particular day.*

Consistent with the patterns observed in most F2P games, many users never spend money on the game. For example, among customers who churned sometime during their first 28 days of play after installation, over 60% did not use any coins (recall that every user starts the game with an endowment of 70 coins) and over 99% did not spend any money. Among the users who survived their first 28 days (i.e., we observe them playing afterwards), 64% of them used at least one (free) coin and 4.1% spent money during their 28 first days. While this difference in monetization rates between early churners and user who keep playing is partially driven by the increasing complexity of the game at later stages, it also highlights the importance of retaining users to achieve high levels of in-game expenditure.

We now explore how some of the specifics of the game design affect users' engagement, retention and monetization. Note that these model-free analyses are not meant to precisely estimate the impact of game characteristics and game progress on customer behavior (to do so we would need a clean and exogenous measure of progress). Rather, we aim to present evidence in our data for the link between the game design and progress, and its relationship with user engagement, retention, and monetization; which we later leverage with data from a large-scale field experiment.

3.2.2 Progress and game design

As with most puzzle-type games, the main motivation to play the game is to pass new levels and keep making progress within the game. Accordingly, the ability to make progress affects users' retention, engagement, and monetization. Figure 2 shows the relationship between the progress achieved — measured by number of new levels passed on one day of play — and the likelihood that the user will play again on the following day. There is a clear positive relationship between these two phenomena; users are more likely to return and play if they have experienced more progress in the game in the previous day. Furthermore, conditional on having lives to continue playing, users are 14% more likely to stop playing for at least 10 minutes after failing to pass a level in the current round than when progress was made (i.e., right after passing a new level for the first time).

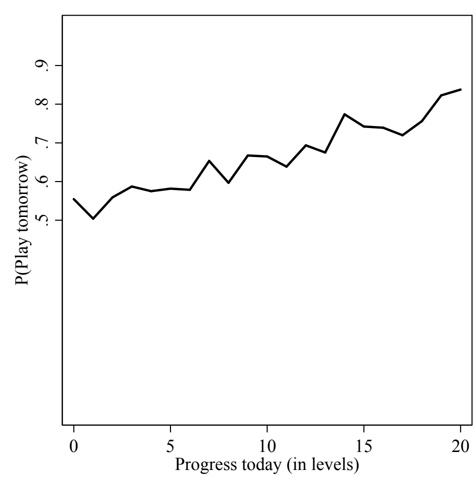


Figure 2: Probability of playing again tomorrow as a function of progress today. *Progress today is measured as the number of new levels that the user passed in a particular day (levels that the user did not pass in the past).*

Despite the benefits of making game progress on users' engagement and retention, game designers introduce mechanisms that purposefully delay progress in order to create a sense of challenge and to monetize the game. Not surprisingly, these mechanisms affect user retention and engagement. One mechanism that directly prevents customers from making progress is the use of gates, where users need to wait for 5 days before being able to play higher levels, unless they use coins to break the gate. In Figure 3-left we can see that users are more likely to churn (defined as not playing for at least 30 days) after passing a gate level—i.e., levels 40 and 60—than in non-gate levels. Moreover, we observe that users are more likely to stop playing right after passing a "gate" level (recall that after passing a gate users can still play lower levels of the game), than after passing a non-gate level (Figure 3-right), where "stop playing" is defined as not playing another round for at least 10 minutes, conditional on having lives to continue playing. These figures provide model-free evidence that users are less interested to continue playing when they know that they cannot make any progress in the game, despite having passed the last level and having lives in their stock.

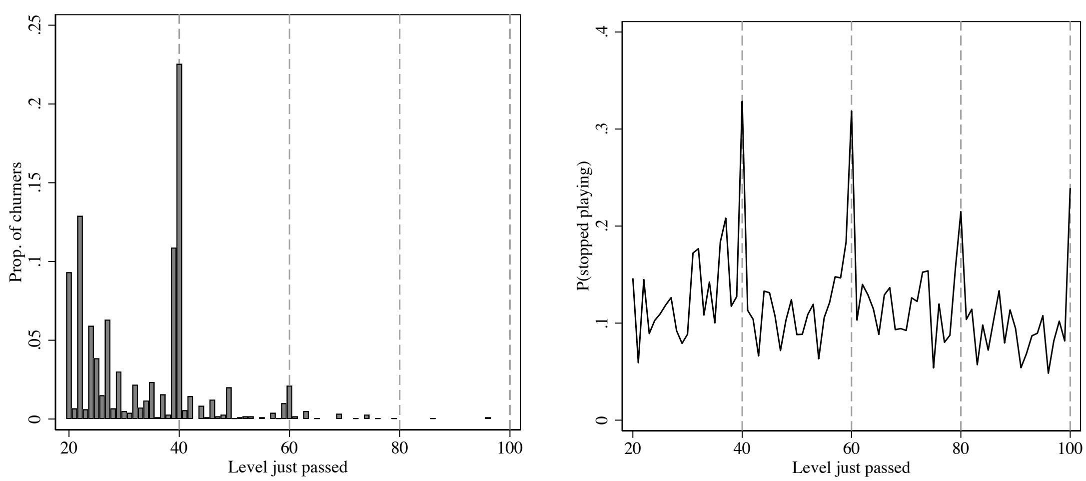


Figure 3: Retention and engagement when stopped at gates. *The figure on the left shows a histogram of highest level passed by churners (defined as users who have not played in at least 30 days). The figure on the right shows the proportion of users who stop playing (i.e., the user does not play for at least 10 minutes, even though they had lives available) right after passing a level for the first time. Vertical dashed lines mark "gate" levels. Only consider users who passed level 20 are considered.*

One may wonder why the firm would introduce gates given their higher churn rate. Gates are a place where the free part of the game "meets" the firms' monetization mechanism, enticing customers to spend money. Indeed, one of the most common uses for coins is to allow users to break a gate and therefore be able to play higher levels of the game without waiting the five days. We corroborate this behavior in Figure 4 where we show the average propensity to spend money by the highest level achieved in the game. The figure shows that users are indeed likely to use coins right after passing a gate level, moment at which they can break a gate with coins.

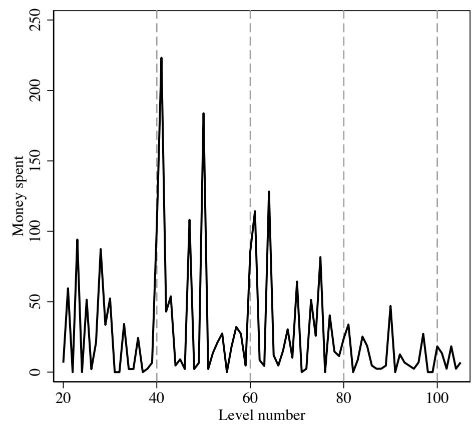


Figure 4: Total money spent by level. *Total amount of money (multiplied by an unannounced factor to disguise firm's revenues) collected in each level of the game. Vertical dashed lines mark "gate" levels. The pattern shows clear peaks right after a user has passed a "gate" level (e.g., levels 40, 60, 80), indicating that the game obtains meaningful revenue from monetizing the gates. Note that this analysis does not control for the number of users playing each level.*

To investigate the relationship between game progress and monetization, we focus on the observations when the users are stopped at a gate, and relate it to the progress rate of the user prior to arriving to the gate. Figure 5 shows the average number of new levels passed (denoted as progress) on the date of reaching the gate, split by whether users spend money or coins at a gate. Consistently across both figures, users who have made more progress during the day are more willing to spend money or coins to pass that gate without the need to wait for five days.[14] Thus, we find that users' past progress affects the willingness to spend money and coins on a gate. That is, users have stronger willingness to continue playing higher levels—and therefore are more

[14] The correlation between progress made and the likelihood to use coins in the gate is 0.190 (p-value = 0.000) and the correlation between progress made on that day and the likelihood to spend money to break the gate is 0.131 (p-value=0.0001).

willing to pay for it—when they have made greater progress during that day. This may suggest that helping users make progress (possibly via reducing game difficulty) can not only lower their likelihood of churning at the gate but also paying to pass it.

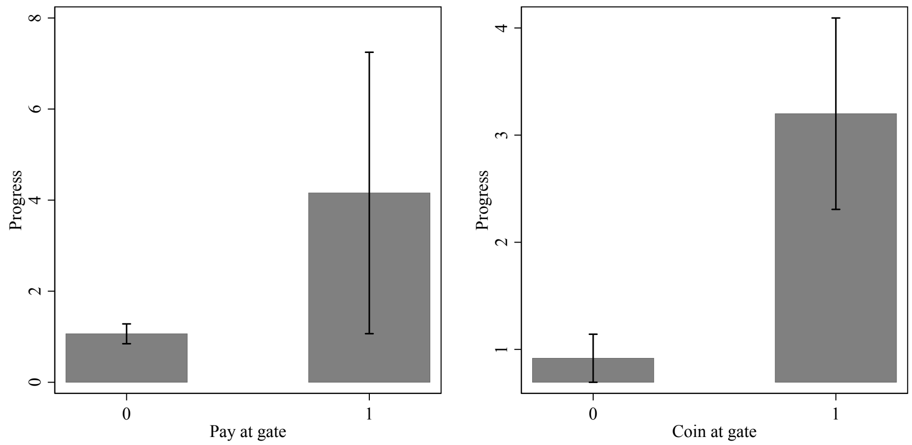


Figure 5: Monetization in gates and progress made. *Progress is measured as the number of new levels that a user has passed on that playday. The progress (number of new levels played) is significantly higher for users who paid at the gate relative to user who did not (left-figure; t=–3.82, p-val<0.001) and for users who used coins at the gate relative to user who did not (right-figure; t=–5.63, p-val<0.001). Only users stopped at a gate are considered in these figures.*

3.2.3 The role of game difficulty

Another mechanism that directly affects the extent to which users can make progress in the game is the difficulty of each round. The overall difficulty of each round is determined by the combination of two factors: (1) the *goal* of the level being played which is fixed for each level (e.g., some levels require users to achieve at least a certain number of points and others to achieve a certain number of piece combinations within a time period) and (2) the *allocation of the pieces* on the board. As described earlier, a "move" in this game involves combining three or more pieces of the same color. If the game designer includes more pieces of the same color together, achieving the goal is easier. Similarly, there are "special chips" that allow the user to eliminate multiple pieces at once, also increasing the score and the chances to pass a level. While the exact location of each piece is random — pieces are shuffled in each round — the probability of finding two or more pieces of the same color is consistent across all rounds of the same level. The same is true for the occurrence of "special chips."

The game design includes a wide range of difficulty between levels. While there is an overall trend of increasing difficulty as the game progresses, the level of difficulty can vary dramatically from one level to the next. Some levels are very easy to pass, requiring on average two to three rounds (attempts), whereas other levels are much more difficult and tend to require many more rounds (sometimes 20 or more) before users pass them successfully. In fact, many blogs and forums are dedicated to discussion of notoriously difficult-to-pass levels in popular games.[15] Consistent with the relationship observed between game progress and user behavior, we observe a strong relationship between the difficulty of a level and the retention and monetization of users who are at that level. We measure the difficulty of a level as the number of rounds that users need to play, on average, before they pass a level. We then compute the correlations between the difficulty metric and monetization and churn measures. To have a consistent metric of difficulty across levels that controls for survival bias (i.e., weaker players are more likely to abandon the game earlier on and therefore would be over-represented in early levels), we only consider users who (ex-post) achieved level 100 when calculating game difficulty.

[15] https://thecandycrush.com/6-insanely-hard-candy-crush-levels/

	P(coin)	P(money)	% users who churn	P(churn)
Correlation with difficulty	**0.302**	**0.553**	**0.880**	**0.394**
P-value	(0.0001)	(0.0018)	(0.0000)	(0.0106)
# obs (game levels)	100	100	100	100

***Table 1: Correlation between level difficulty and user behavior**. % users who churn is the proportion of users who ever reached a particular level but did not make it to the next (during our observation window of at least 30 days). P(churn) is the propensity to churn in each round played. We include levels 1--100 as they have a large number of observations. Results are robust to exclude gate levels, which would avoid the possible confound with the use of coins/money for passing the gate.)*

As can be seen in Table 1, the probability that a user will use coins or spend money when playing a round in a particular level is positively correlated with the difficulty of that level (0.302 and 0.553, respectively). Similarly, difficulty is strongly correlated with customer churn, which we measure in two ways. First, we compute the proportion of customers who achieve a level but churned (did not play in the following 30 days). The correlation between that metric and level difficulty is extremely high (0.880). This metric is a bit complicit with game difficulty because users play many more rounds in high difficulty levels. Therefore, we also compute the proportion of rounds played at a particular level that were the "last round ever played" by users (with "last round" being determined by the use not playing again for at least 30 days). We also find a positive correlation between the probability of churn in a particular round and the difficulty of each level (0.394), "controlling" for the number of rounds the player played at that last level.

Overall, our model-free analyses highlight the tradeoff between short-term benefit to the firm due to IAPs and long-term benefit due to higher retention. Specifically, gates and increased game difficulty both increase short-term monetary spending, but at the same time decrease game engagement and increase churn. We also demonstrate that the link between game difficulty and churn is, at least partially, related to the ability of consumers to make progress; when consumers fail to progress (e.g., due to difficulty or a gate), they are more likely to churn. Furthermore, exploratory analyses show that even in the short-term, decreasing game difficulty may increase game monetization at gates because consumer are more likely to pay for gates when they make good progress prior to reaching the gate.

In the remainder of this paper, we pay special attention to the role of difficulty (and less so to the role of gates) because of its direct link to game progress but also because it is a product characteristic that can be easily personalized to one player without altering the shared experience for the overall player community. Difficulty can also be manipulated within an individual user (e.g., it can be adjusted differently across playing occasions), allowing the firm to intervene in a dynamic manner when its impact is expected to be highest and/or to prevent impending churn. As a result, it can serve as a very powerful instrument to exogenously manipulate the extent to which a customer can make progress in the game, as we describe next.

3.3 Field experiment

Building on our model-free analyses, we want to empirically and causally investigate the tradeoffs and interdependencies between retention and monetization in freemium settings. Specifically, we want to understand the short- and long-term effects of product design on customer retention and monetization to provide guidance to firms in their efforts managing customer retention and monetization through the design of freemium products/services. We leverage a field experiment that allows us to investigate this question by manipulating the users' ability to make progress in the game. In particular, we obtain data where the company exogenously manipulated the difficulty that each individual user faces—thus affecting their likelihood to make progress in the game. Importantly, difficulty is manipulated at the user level, leaving the rest of the game design intact. That is, users play the exact same levels as the ones originally designed in the game, in the same order of appearance, the only difference being that some users face more favorable piece allocations on the board.

As described in Section 3.2.2, the game was originally designed such that pieces are shuffled on the board with the same probabilities of color co-occurrences across users facing the same level of the game. However, in 2014, the collaborating company implemented a new technology that allowed them to manipulate the baseline level of difficulty on a user-per-day basis. Specifically, at the start of the day, a sample of users could be assigned a different difficulty level, whereby the chances of getting adjacent same-color pieces as well as "special chips" were higher than the regular rate. The company implemented such a technology with the intention to reduce churn among mature players.[16] To do so, they adopted the following policy: if a user had played less than 20 rounds in the last seven days, their difficulty would be decreased, increasing the chances of color co-occurrence by 20% and for special chips by 10%; if the user had played less than 15 rounds in the last seven days, the difficulty would be further decreased, and so forth. If a user had played less than five rounds, or had not played at all in the last seven days, the difficulty would be set at its lowest level, with the chances of color co-occurrence increased by 100% and the chances of obtaining special chips increased by 45%. (See Appendix A1 for further details on the intensity of the manipulation and a visualization of the game.) Importantly, the company implemented this new policy in a controlled randomized experiment (A/B test), leaving half of the qualifying players "untouched" over the course of 50 days. Specifically, starting June 11th, 2014 every user who qualifies to be part of the experiment — i.e., a player who has already passed level 20 and whose number of rounds in the past 7 days is lower than 20 — would be assigned to the *lower difficulty* condition (i.e., the treatment) with a probability of .5, and would be assigned to the *default difficulty* condition (i.e., the control) otherwise. This setting results in a very

[16] For the focal company, a "mature" player was a player who had already passed level 20, who had already shown a certain level of commitment for the game.

favorable scenario in which we observe players with the same characteristics, same past play, who are trying to pass the same level but face different degrees of difficulty.

Note that the randomization was performed at the player level, that is, once a user is allocated to treatment or control, the user will stay in that condition throughout the duration of the experiment (until August 3rd, 2014). However, being assigned to the treatment condition does not imply that the user would face lower difficulty every day of play. Rather, at midnight of every day (specifically, at 00:00 UTC) the number of rounds the players played in the past seven days is assessed, and players in the treatment condition receive reduced game difficulty only if they played fewer than 20 rounds in the past seven days and according to the number of rounds they played. The difficulty then stays fixed for the next 24 hours. This characteristic of the experiment has important implications for our analysis.

Similar to other longitudinal experiments where randomization is done cross-sectionally (in our case, at the user-level) the observations for days 2, 3, etc. (and even rounds 2, 3 of the first day) after the first intervention should be analyzed with caution as the game difficulty in those periods is no longer perfectly random relative to the control users.[17] This is not to say that we cannot estimate the causal long-term effect of the treatment, which we can easily do by comparing the behaviors of treated and control users, but we need to exert caution when estimating the causal effect of progress in the game on user retention and monetization when using observations other than the first round.

[17] This is due to the fact that being treated on day 1 might affect the behavior of the user and therefore the conditions they face on days 2, 3, and so on. For example, if being treated increases retention among users, more users in the treatment group will play on day 2 (than those in the control) therefore removing the perfect balance created by the random allocation on day 1. This is also true for the rounds played during the first day; because the treatment in the very first round affected the outcome of that round and likely the ability for the user to make progress.

3.4 Data

Our dataset includes the complete gameplay history (prior to August 3rd) of the 330,000 players who qualified for the experiment. These are the users who had already passed level 20 of the game, and who at some point between June 11th and August 3rd (duration of the experiment), during the previous seven days played fewer than 20 rounds. The level of play during the last seven days is used as a proxy to identify users who are at possible risk of churn, as low activity is often associated with higher risk of churn.[18] Users joined the game between May 1st and July 3rd implying that we observe every user from the moment they played their very first round in the game, and for at least one month after that. During this time, users play a total of 79 million rounds, from which we observe the following characteristics:

- Level played: Denoting the Level the customer is playing on that round
- Outcome of the round:
    - `Win`: did the player pass the round?
    - `Stars`: # stars (from 0 to 3) obtained in the round
    - `Points`: numerical score obtained in the round
    - `Combination size`: maximum number of pieces that were connected to form a snake.
    - `Moves`: number of cell combinations formed in total
    - `Time`: # seconds that the round lasted
- Other behaviors of interest:
    - `Coins`: did the player use any coins before finishing the round?
    - `Extras`: did the player use any extra (to make the board more favorable) during this round?

[18] Note that given the very low retention figures in these kinds of services, this "at risk" condition is not very restrictive. Specifically, when we look at existing users who had passed level 20 before the day of the experiment and who are active (i.e., we see them playing at some point after that day), 47.6% of them belong to the "at risk" condition at the moment of the experiment.

- `Purchase`: did the player spend real money to buy game currency before finishing the round?

Table 2 shows the descriptive statistics across the 79 million rounds.

	Mean	SD	p5	p25	p50	p75	p95
Level	34.0	22.8	7.0	17.0	29.0	43.0	77.0
Win	0.39	0.49	0	0	0	1	1
Stars	1.10	0.79	0	1	1	1	3
Points	35,006	19,280	11,650	22,450	31,400	43,050	69,900
Combination size	5.14	1.18	4	4	5	6	7
Moves	15.63	6.94	6	10	16	19	28
Time	121.1	71.5	47	76	106	146	248
Coins	0.007	0.081	0	0	0	0	0
Extras	0.028	0.204	0	0	0	0	0
Purchase	0.001	0.033	0	0	0	0	0

Table 2: Descriptive statistics of round-level data. *N = 79,030,000 rounds.*

We also create a set of metrics capturing user heterogeneity in their skills and level of engagement prior to treatment. Those metrics are based on the users' activity prior to passing level 20—as all qualifying users were required to have passed that level. We define `Level20` variables as the number of days/rounds played before level 20, total number of stars and coins collected before level 20, and whether the user had used coins/extras before level 20. We also use the round-level panel data to capture user-level variables that change over time and that will be relevant for the analysis. These include `Age` variables (e.g., maximum level achieved, tenure with game), `RFM` variables (e.g., amount of play, days since last play), `Stuck` metrics (e.g., # rounds in the current level, proportion of wins in the last day), and `Skill` variables (e.g., average # rounds per level, average # stars per level). Table 2 shows the summary statistics for the most relevant variables (see Appendix A2 for the full list of variables).

		Mean	SD	p5	p25	p50	p75	p95
`Level20`	# rounds	45.83	36.91	23	29	36	49	99
	Did use coin	0.513	0.5	0	0	1	1	1
`Age`	# rounds	189.6	217.7	35	62	113	225	615
	Max level achieved	37.74	16.97	20	24	39	40	72
`RFM`	# days since last play	13.88	17.9	1	3	7	16	54
	# rounds last week	5.728	6.68	0	0	2	12	18
`Stuck`	# rounds in this level	26.72	62.45	0	2	7	25	115
	# playdays in this level	3.054	3.804	0	1	2	4	10
	Prop. wins yesterday	0.308	0.341	0	0	0	1	1
`Skill`	Avg. # rounds/level	4.410	3.653	2	2	3	5	11
	Avg. # stars/level	1.988	0.257	2	2	2	2	2

Table 3: Descriptive statistics for the user-level variables*. Level20 variables were measured at the moment the customer passed level 20. The rest of the variables are dynamic (i.e., values change over time) and are summarized using the values on the day each user qualified for the experiment. N = 329,999 users.*

There is a rich variation across users. For example, it takes users an average of 45.8 rounds to pass level 20, with a standard deviation of 36.9, and 51.3% of the players use some coins even before passing level 20. Their level of play at the moment of the intervention is very heterogeneous as well; users play an average of 189.6 rounds before being treated (or assigned to control), with the maximum level achieved ranging from 20 (5 percentile) to 72 (95 percentile). There is also rich variation in how stuck players are — the number of rounds in the current level ranging from 0 up to 115 — and also in their skill levels — the average number of rounds per level is 4.4, with that figure ranging from 2 up to 11. This variation offers an opportunity to explore the value of reducing game difficulty for different players.

4 ANALYSES AND RESULTS

4.1 Randomization and manipulation tests

We first confirm that the randomization was well executed by comparing the distributions of the user-level variables at the moment of the intervention. We verify that there are no systematic differences across conditions (see Appendix A3 for full set of results). Second, we corroborate that the intervention caused the intended effect of making the game easier for treated users. To do so, we look at four outcomes that are directly affected by the difficulty of the round namely `Win`, `Stars`, `Points`, and `CombinationSize` at the observations (rounds) of treatment and control players in the first day of the intervention. Specifically, we run a linear model for each outcome against a treatment dummy and cluster the standard errors at the user level (unit of randomization). See Table 4 for the results. All outcomes show a substantial and (statistically significant) positive change for users in the treatment condition; users have a higher chance to win, collect more stars, get more points, and are able to create longer snakes. For example, users in the treatment condition on average earn 7,382 points and 0.297 star more than those in the control condition. All these outcomes were expected as treated customers faced more favorable board allocations.

	Win	Stars	Points	Combination Size
Treatment	**0.141**	**0.297**	**7,382**	**0.172**
	(0.002)	(0.003)	(64.7)	(0.004)
Constant	0.368	1.048	35,412	5.157
	(0.001)	(0.002)	(36)	(0.002)
# obs	2,009,966	2,009,966	2,009,966	2,009,966

Table 4: Manipulation checks. *OLS of the round outcome against a treatment dummy using all rounds on the first day of the experiment. Standard errors are clustered at the user level. Bold numbers indicate that p-value<0.01.*

We also run separate regressions for each level of difficulty reduction—recall that the intensity of treatment changed by the amount of play in the previous week. As expected, the impact of the treatment on these outcomes is stronger for user with lower playing level in the seven days prior to treatment (more pronounced difficulty decrease). For example, looking at the effect of treatment on the probability to win and on the number of stars collected in each round (Table 5), we observe that the magnitude of the effect increases monotonically as the treatment gets more severe.[19] We therefore conclude that the randomization was well executed and was successful at manipulating game difficulty as intended.

	Win (difficulty 5 = default)				Stars (difficulty 5 = default)			
	4	3	2	1	4	3	2	1
Treatment	**0.030**	**0.076**	**0.133**	**0.190**	**0.052**	**0.135**	**0.249**	**0.419**
	(0.004)	(0.004)	(0.004)	(0.002)	(0.005)	(0.006)	(0.006)	(0.003)
Constant	0.327	0.347	0.358	0.393	0.983	1.012	1.030	1.090
	(0.002)	(0.002)	(0.003)	(0.001)	(0.003)	(0.004)	(0.004)	(0.002)
# obs	370,137	302,836	281,827	1,055,166	370,137	302,836	281,827	1,055,166

Table 5: Manipulation checks by degree of difficulty. *OLS of Win and Stars against a treatment dummy using all rounds on the first day of the experiment. Standard errors are clustered at the user level. Treatment variable in bold indicates p-value<0.01.*

4.2 Evaluating the impact of the intervention on player behavior

We turn to explore the impact of the intervention on customer behavior, in particular, on the behavioral outcomes that are most relevant to understanding retention and monetization in freemium settings. We compute, for each player, the cumulative number of playdays, rounds, money expenditure, and use of coins for each day after the user was treated (or assigned to control). As Figure 6 shows, customer behavior significantly changed due to the intervention. As expected, the amount of play increases among treated users, and such increments get larger as time goes by. For example, within a month of being treated (or assigned to control), treated users have played, on average, one extra day and 10 rounds more than control users have, and have also made more progress in the game, achieving level 44 instead of level 42. The magnitude of these effects is meaningful for the focal firm as those increments translate to approximately 20% increase in overall customer play.

The intervention not only increases gameplay and engagement, but it also has a positive impact on monetization, with treated users spending more money and more coins than those in

[19] This pattern is consistent across outcomes (results for `Points` and `CombinationSize` are shown in Appendix A3). In the remainder of the analyses, we report the average treatment effect, i.e., across the four treatment intensities.

control (see last two charts in Figure 6). Not only does this finding go against the company's prior expectations, but it seems at odds with the preliminary evidence (Section 2) where we show that users tend to spend money and coins when progress is prevented (gates and difficult levels). This difference is already visible on the first day of the intervention and increases monotonically over time. Thus, in the tradeoff between offering more of the free product (easier game) and encouraging consumers to purchase in-app goods, we find that by providing more of the free game (easier game), the firms increases, rather than decreases, long-term IAPs.

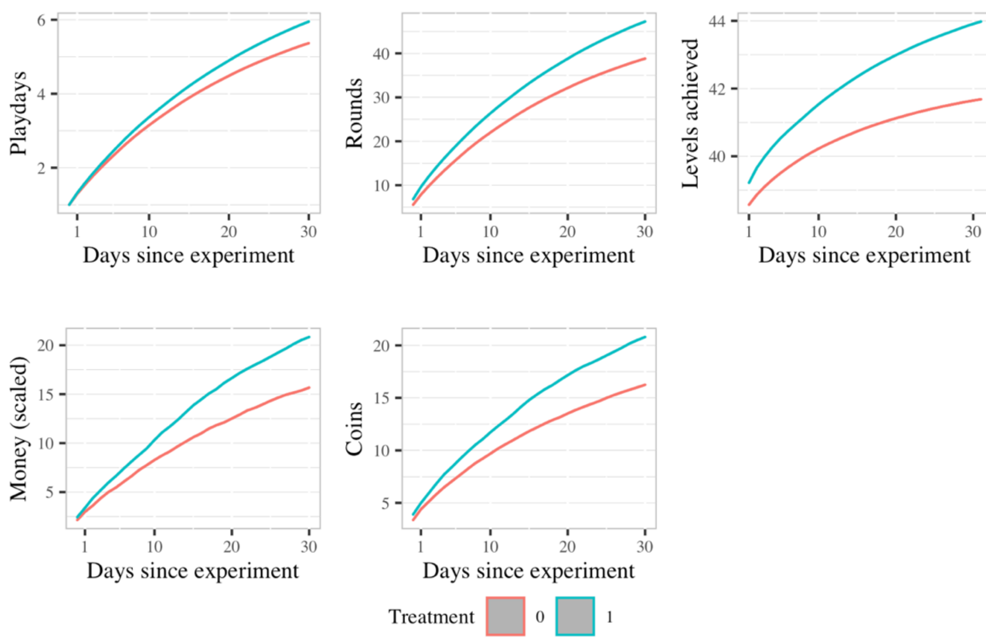


Figure 6: Cumulative behavior across experimental conditions. *These figures show the cumulative number of playdays, rounds, money spent (scaled by an unknown amount) and coins used the game. Standard errors around the lines are reflected in the graphs but are not detectable due to their small magnitude.*

Nevertheless, from this aggregate analysis we cannot conclude that a reduction in game difficulty (or the increase in the ability to make more progress in the game) *directly* affects the user tendency to spend coins and money. This is because lower difficulty makes it less likely to run out of lives, and therefore allows the user to play more rounds and make more progress, having the chance to use more extras, coins, etc. Additionally, if the ability to make progress in the game increases retention and engagement — as we expect given the preliminary evidence — treated users will be more likely to play again after being treated than control users. Therefore, the increase in the use of money or coins might not be a direct impact of the ability/inability to make progress but just an artifact of treated users playing more.

Furthermore, the experimental design was such that whether a treated user receives difficulty level 1, 2, 3, 4, or 5 depends on the number of rounds they have played in the previous seven days (recall that control users always get difficulty 5). Therefore, it is possible that treated users get different intensities of treatment when they play the game after the first day in which they were treated.[20] This means that while the allocation to the treatment group is fully randomized, the treatment intensity, or the degree to which the ability to make progress is manipulated, is not constant over time, and hence can be endogenous *after* the first day of the experiment.

While these idiosyncrasies do not prevent us from making causal statements about the overall impact of the intervention (the intervention *causing* the differences in behavior presented in Figure 6), we cannot (yet) make causal statements about the direct impact of *progress* on user retention and monetization in a specific day or a round. To do so, we need to analyze the data in a way that controls for the users' ability and propensity to play as well as for the treatment intensity.

4.2.1 Impact of treatment on user engagement and retention

We analyze player behavior on the first day of the experiment, i.e., the moment at which the randomization took place. Table 6 shows the results from a linear regression of engagement and retention outcomes on treatment. (Using this intent-to-treat formulation implies that, across all regressions, the coefficient for the variable treatment is an unbiased comparison between treated

---

[20] See Appendix A1 for an illustration of how treatment intensity may vary over time depending on the user amount of play.

and control users.) Engagement is computed at the end of each round and is measured as whether a user decides to play another round within the next 10 minutes conditional on having lives (i.e., we discard the observations when users run out of lives). Retention 1 is measured as whether a user plays again tomorrow, Retention 7 as whether the user plays at least once within the next 7 days, and so forth. We find that treated users play more rounds and make more progress than the users whose difficulty level was unchanged (on average, 1.26 additional rounds played and 0.746 additional levels achieved). Treated users are also more likely to continue playing than those in the control. While the last two metrics might be driven, in part, by whether a player has lives left to continue playing, the first metric, which we denote "engagement", controls for whether a user has the ability to continue playing for free. Therefore, it provides very robust evidence that lowering the difficulty of the game, enabling users to make more progress, significantly increased engagement.

	# Rounds played	Progress made	Engagement	Retention 1	Retention 7	Retention 14
Treatment	**1.2470**	**0.7460**	**0.0213**	**0.0270**	**0.0246**	**0.0199**
	(0.0251)	(0.0088)	(0.0007)	(0.0016)	(0.0017)	(0.0016)
p-val	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Constant	5.5650	0.7840	0.8180	0.2950	0.6430	0.7430
	(0.0132)	(0.0039)	(0.0005)	(0.0011)	(0.0011)	(0.001)
# obs	329,999	329,999	1,867,849	326,472	316,257	308,003

Table 6: Impact of the intervention on daily engagement and retention outcomes. *OLS of the behavior of interest (on the first day of the intervention) against a treatment dummy. The number of observations for "Engagement" corresponds to the rounds in which a user has lives to continue playing. Standard errors are clustered at the user level. All other regressions are at the individual level. The number of observations drops for the retention metrics because some users were treated for the first time just a few days before the end of the observation period and therefore were not observed for a long enough window. Robust standard errors (in parenthesis) and p-values are reported. Treatment variable in bold indicates that p-value<0.01.*

Similarly, when looking at the proportion of users who play again the day after (Retention 1), 7 days later (Retention 7), and 14 days later (Retention 14), we find that the treatment significantly increased retention among users.[21] As discussed earlier, this significant increase in retention rates might have caused the positive impact on monetization rates (Figure 6). That is, treated users generate more expenditures in the long-run not because greater progress directly increases monetization, but rather because progress delays churn, giving treated users more opportunities to spend money in the game.

4.2.2 Impact of treatment on user monetization

To separate the direct effect of the intervention on customer spend from that of increased retention (due to the intervention), we analyze the consumption of extras and IAPs at two different levels: *Per-day monetization*, looking at the total amount of rounds in which a user consumes extras, coins and money on their first day (these figures correspond to the left-most observation of each of the lines presented in Figure 6), and *per-engagement monetization*, measured as the propensity to use premium content in a particular round. Because the later metric is conditional on users playing that particular round, any effect observed on monetization fully "controls" for engagement or retention differences across users.[22] That is, this metric captures the *direct* effect of the treatment on the propensity to spend in the game, free from any change in expenditure due to increased engagement. Table 7 shows the results from the regression analyses.

Indeed, treated players use more extras, use more coins, and spend more money (than control users) already on the first day in which they are treated. However, the opposite is true at the more disaggregate level, when we remove the effect of engagement/retention: Conditional on playing a round, the probability of using extras or coins in that round is lower for customers under the

---

[21] We replicate the analysis using retention in the next 3 days, 28 days, and 45 days, obtaining similar results for all metrics, with diminishing effects as the retention window is longer.
[22] One might still worry about potential bias induced by the type of users who continue playing; in Section 4.3 we discuss that possibility and present analyses testing the robustness of our results to that potential concern.

treatment condition. This negative effect of game difficulty on extras and use of coins, is not surprising as using extras is equivalent to making the game easier (the user has higher chances to combine pieces on the board) and therefore, when the game is easier by design, the user would face lower need of extras.

	Per-day monetization			*Per-engagement monetization*		
	# Rounds with extras	# Rounds with coins	# Rounds with money	Prob(Use extras)	Prob(Use coins)	Prob(Use money)
Treatment	**0.0075**	**0.0051**	**0.0012**	**–0.0025**	**–0.0004**	0.0000
	(0.0010)	(0.0007)	(0.0003)	(0.0003)	(0.0002)	(0.0001)
p-val	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0051*	*0.6340*
Constant	0.0922	0.0390	0.0057	0.0229	0.0078	0.0013
	(0.0007)	(0.0004)	(0.0002)	(0.0002)	(0.0001)	(0.0000)
# obs	329,999	329,999	329,999	2,009,966	2,009,966	2,009,966

Table 7: Impact of the intervention on monetization. *OLS of the behavior of interest (on the first day of the intervention) against a treatment dummy. For per-day metrics, the number of observations correspond to the number of users participating in the experiment. Robust standard errors are reported in parentheses. For per-engagement metrics, the number of observations corresponds to the rounds used for each regression. Standard errors (in parentheses) are clustered at the user level. Treatment variable in bold indicates that p-value<0.01.*

However, surprisingly, we find that even in the short-term (first day of intervention) this effect turns positive due to increased engagement and retention, putting in doubt the firm strategy to increase difficulty to generate short-term spending (though the probability to use money is not significantly affected by the game difficulty). While these figures are small in magnitude, their effect is very meaningful to the firm given that those behaviors are quite rare in practice. To better understand the monetization behavior, we further separate per-engagement expenditure — both coins and money — by whether the user spends the currency on extras (to pass levels) or on gates (to break gates). Table 8 shows the results from this analysis.

Another interesting finding emerges, lowering the difficulty of the game significantly *increases the propensity to spend* coins and real money *on gates*. Note that the gate-related dependent variables (coin gates and money gates) are conditional on users being stopped at a gate—i.e., this effect is not due to the fact that users are more likely to get to gates—and therefore this treatment effect is only attributable to a higher propensity to use coins and money, once at a gate. This effect is interesting because it highlights that offering more of the free product (which a priori is expected to reduce purchases of the premium features) leads to the opposite effect and entices users to spend more.

	P(Coin extra)	P(Coin gate)	P(Money extra)	P(Money gate)
Treatment	**–0.00059**	**0.00082**	–0.00007	**0.00019**
	(0.00012)	(0.0002)	(0.00005)	(0.00008)
p-val	*0.00000*	*0.00003*	*0.13000*	*0.02540*
Constant	0.00449	0.00548	0.00069	0.00105
	(0.00449)	(0.00548)	(0.00069)	(0.00105)
# obs	2,009,966	652,574	2,009,966	652,574

Table 8: Impact of the intervention on the type of expenditure. *OLS of the behavior of interest against a treatment dummy. Standard errors (in parentheses) are clustered at the user level. The number of observations corresponds to the rounds used for each regression. The gate-related outcomes are conditioned on the user being at a gate Treatment variable in bold indicates that p-value<0.03.*

While the main focus of the literature on product design in freemium services has been on the notion that "enhancing" the core product (or increasing the portion of the product that is free) would *increase* retention at the cost of *reducing* monetization. Here we show evidence that when looking at customers at risk of churning (about 50% of the users at any point in time), the tradeoff no longer exists; in turn, there is a positive synergy between retention and monetization whereby altering the product design to give away more for free can increase engagement and retention of users while increasing the chances that they will spend money on the service both in the short-term (the first day of the treatment) and in the long-term (as can be seen in Figure 6).

4.3 Robustness checks

One possible caveat to the round-level analysis (Table 7 and Table *8*), is that the users who continue playing longer — and therefore are more represented in this analysis — might be systematically different between the treatment and control groups. This is a form of survival bias in which users who kept playing after being treated with an easier game in the first round (or first few rounds) are different from the control condition users who kept playing after being faced with higher difficulty levels. We run two robustness checks to address this concern. First, we replicate the round-level analyses using the first five rounds of each player after treatment.[23] We use the first five rounds as a conservative measure for the number of rounds that most users would play regardless of their treatment allocation. Because all users are endowed with five lives at the beginning of each day, even those who face the highest difficulty level can play up to five rounds without the need to wait or to buy any extras. Second, we replicate all analyses controlling for all the user characteristics described in Section 3.4. Recall that these characteristics capture users' prior skills in the game, levels of activity, and previous propensity to use extras, coins and money. See Appendix A4 for the results from the robustness tests. Briefly, (with the exception of coins and money used on extras within the first five rounds where significant differences are not found), all other effects are consistent with the results presented in Table 7 and Table 8.

4.4 Proposed Mechanism

Summarize the results so far, we find that users who experienced lower game difficulty continued to play longer relative to control users, above and beyond the mere effect of having lives. This effect is consistent with both the model-free evidence and the finding that treatment increases the chances that a user will play again the day after. Regarding monetization, the intervention of lowering game difficulty was also beneficial for the firm as it increased the use of extras, coins, and in-app spending. However, analyzing behavior at the per-round level depicts a different pattern than the one obtained when analyzing aggregate outcomes. Once we control for

[23] One way to overcome this concern entirely would be to replicate the analysis using only the very first round of data. However, doing so discards a lot of useful information, dramatically reducing the sample size and thus the power to capture the observed effects.

ability to play — in our case, by modeling outcomes conditional on play — we find that treated users are *less likely* to spend coins and money on extras but are *more likely* to spend in gates. All in all, we find evidence of two types of positive synergies from engagement/retention to monetization. First, even though reducing the difficulty of the game decreases the propensity to spend on extras, the impact of such a change on engagement and retention is sufficiently strong that the net effect becomes positive merely from engagement/retention. This net positive effect of game difficulty on monetization holds true even in the short-term (first day of intervention). Second, the increase in retention due to reduced difficulty is associated with an increase in the IAPs for goods that can help them to continue playing the game — in this case, gates — both in the short- and the long-term.

Collectively, these findings are consistent with the notion that making progress in the game increases the retention and engagement of players, while affecting their propensity to spend. While users are less prone to spend currency on making the game easier when difficulty is decreased, they are more likely to spend currency on eliminating the other barriers (i.e., gates) that prevent users from continuing to make progress. To more directly investigate the role of game progress as one of the underlying mechanisms for the increase in engagement, retention and monetization we run an instrumental variable (IV) regression with the behaviors of interest (engagement, retention and monetization) as dependent variables but now include "progress" as an independent variable, instrumented by the treatment allocation. Specifically, we measure game progress at the end of each round, defined as the proportion of new levels a user had completed successfully (so far) on that particular day, and relate that variable with the propensity to continue playing, propensity to use extras, coins, etc.

Game progress is not completely exogenous as it depends on several other decisions made by the user on that day (e.g., the decision to continue playing) and therefore the coefficient of game progress in a linear regression of game progress on customer behavior is likely to be biased. We obtain unbiased estimates of the causal effect of progress on behaviors by instrumenting the progress variable with the treatment allocation. Treatment is a valid instrument for these analyses as it was randomly allocated across users and is strong because the treatment allocation significantly increased the chances to make progress in the game. Table 9 presents the results of the IV analysis, confirming that progress has a causal effect on the behaviors of interest. We replicate these analyses using different proxies for progress, namely "whether the user made progress (passed a level) in the most recent round" and "the ratio between the progress made so far today (level passed today) and the average daily progress made by the user prior to the intervention." The results are all consistent with the results in Table 9 and are presented in Appendix A5.

	Next day retention	Continue playing	Use extras	Coin gate	Coin extra	Money gate	Money extra
Progress	**0.1950**	**0.0805**	**-0.0297**	**0.0057**	**-0.0065**	**0.0012**	**-0.0009**
	(0.0139)	(0.0041)	(0.0024)	(0.0004)	(0.001)	(0.0002)	(0.0004)
p-val	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0287
Constant	0.2940	0.8220	0.0402	-0.0007	0.0076	-0.0002	0.0012
	(0.0044)	(0.0016)	(0.0009)	(0.0001)	(0.0004)	(0.0001)	(0.0002)
# obs	188,783	1,191,700	1,116,023	1,116,023	1,116,023	1,116,023	1,116,023

Table 9: Impact of progress on behavioral outcomes*. 2SLS regression of the behavior of interest against a variable that captures the progress made so far. "Progress" is instrumented with the treatment variable. Standard errors (in parentheses) are clustered at the user level. The number of observations corresponds to the rounds used for each regression. Progress variable in bold indicates that p-value<0.01.*

While these results provide convergent evidence that progress is indeed the mechanism driving retention and monetization, it is important to note one possible caveat of the IV analyses. The validity of any IV analysis relies on the assumption that the instrument affects the outcome *only* through the endogenous variable. In our case, this translates to assuming that treatment (i.e., reduction of difficulty) affects the behavioral outcomes only via progress, and not by any other mechanism. While we have strong evidence that the treatment clearly affected progress and other behavioral outcomes, we readily admit that we cannot rule out the possibility that users might also enjoy easier rounds regardless of whether they make progress or not.

4.5 Heterogeneity in treatment effects

In addition to documenting the twofold effect of retention in freemium settings, our results highlight the tradeoffs (or lack of) between retention and monetization, providing actionable insights to firms as to how personalization in the product design can be leveraged to manage customer retention efforts as well as to increase the revenue of each customer. Specifically, the firm can use these insights to identify players who will be most likely to increase revenue — from advertising, premium purchases, or both — via a temporal difficulty reduction.

Arguably, there is no strong need to leverage the heterogeneity in treatment effect in this case because, as it was shown in Section 4.2, the overall impact of the intervention was overwhelmingly positive. Nevertheless, we believe that this exercise is still of interest for two main reasons. First, investigating heterogeneity in treatment effects can further shed light on the account that progress is the underlying mechanism for the observed effect of game difficulty on customer behavior (Section 4.4), if individuals with a higher need or sensitivity for progress exhibit stronger treatment effects. Second, given the ease of product design personalization in online games, as demonstrated by the experimental design of our study, an analysis that identifies individuals with greater benefit from game personalization, illustrates the more general potential value of personalization in the product design, which has been understudied in the marketing literature. The approach taken in this section can be easily replicated by freemium services to guide their targeting decisions in other situations in which the product can be personalized or changed dynamically.

Given the analysis conducted in Section 4.4, we expect that users who tend to seek more progress, or situations in which users seek more progress, will exhibit the strongest treatment effect. We use users' behavior *prior* to the intervention to identify individual players whose long-term value is most likely to increase as a result of a reduction of difficulty (which we call "target" users). We then compare the impact of the intervention on future outcomes of players in those identified groups, relative to individuals with similar characteristics in the control, and compare it to the effect of the intervention on the rest of the sample.

4.5.1 Identifying targets: Heterogeneity in treatment effect

We identify potential targets in two different ways. First, we identify "types" of users *who* would be more responsive to the game difficulty reduction treatment. We do so by capturing intrinsic differences across users like, for example, whether a user is more/less prone to spend money in general. Second, we identify *when* it might be a good moment to intervene. Even for users who might not generally benefit from reduction in difficulty, there might be circumstances (e.g., when they are increasingly frustrated because of lack of progress) that make them more responsive to the intervention.

For the first type of metrics, the "who", we rely on the behavior up to level 20. This is in the spirit of Padilla and Ascarza (2020) who use the first transaction of users to identify customers with higher responsiveness to marketing interventions. We use the behavior up to level 20, in part, because every user in our sample reached that level and therefore everyone is observed up to that point. But most importantly, because it allows for apples-to-apples comparison of the degree of complexity and difficulty of the game across users. If, on the contrary, we used the full history or only the most recent history of each player, users would be playing at very different levels of the game, and therefore the behavior observed would not necessarily reflect intrinsic differences (i.e., customer heterogeneity) but rather differences in the characteristics of the levels being played.

For the second type of metrics, the "when", we use the most recent behavior prior to the moment of first intervention. That way, we can identify characteristics of the current play (e.g., related to the progress made or the need to spend money) that are associated with a stronger sensitivity to the intervention on that moment. This type of information is of great value to developers in the context of gaming and other online/mobile services as they can not only modify the product characteristics individually, but also alter the product features dynamically to better match users' variable needs as they evolve in their use of the product/service.

Based on the findings described in Section 4.4, we identify the following groups of customers:

- *Who:Progress*: Players who seem to enjoy (or seek) progress more than the average player. Given that ability to make progress is driving (at least partially) the results obtained, we hypothesize that the intervention will be more impactful to customers who tend to seek progress in the game. We create two variables to serve as a proxy for progress: *Who:Early_Progress* based on the number of days that the user played the game before reaching level 20. The rationale behind this metric is that users who like progress will tend to progress faster in the game, especially in the early levels, where difficulty is low. We acknowledge that this progress metric may also capture difference in skills of playing the game (as good players are likely to achieve level 20 faster). Accordingly, we create another progress metric.

Who:Progress_Prone capturing how important it is, for a particular user, to make progress in order to continue playing the game. For this metric, we also use all the rounds played before level 20 and compute, for each user, the proportion of times that a user decided to continue playing after having passed a new level (compared to the times that they stopped after having passed a level).[24] Our rationale for such a metric is that users who value progress more will be more likely to continue playing when progress has been achieved.

- *Who:Spender*: Players who are more prone to spend money in the game. While it might seem counterintuitive — one might think the game designers should not reduce difficulty to those who tend to spend in the game, as this will cannibalize sales — we posit that the firm will maximize the return of this intervention by targeting customers who already spent some money. This is because the twofold effect of retention whereby these users will likely increase expenditure due to extended lifetime. Note that we don't expect these users to increase engagement more than the average user, but the mere effect of increasing engagement and retention (like the average customer) will increase their monetization *further* than for the rest of players. We operationalize this with a dummy variable capturing whether the user has made any purchase before level 20.
- *When:Frustrated*: Players who, at the time of first intervention, might be more frustrated than usual, due to lack of progress is recent rounds. We posit that these players need progress more than usual, and therefore their reaction to the intervention might be stronger. We operationalize this variable using the number of rounds that a user has tried (unsuccessfully) the current level (which is the level they were at when first receiving the lower difficulty treatment).

[24] By definition right after passing a level the user has at least one life and hence is eligible to continue playing. Continue playing equals 1 when the user plays another round within 10 minutes of ending the current round.

- *When:Distance-to-gate*: Players who, at the present time, are far away from reaching a gate, and therefore have a higher chance of enjoying progress before getting stopped at a gate. In the main analyses as well as in the model-free evidence, we observe that users are more likely to pay for complementary products (e.g., gates) when they have recently made more progress. Therefore, we posit that the extent to which the intervention increases the purchase for complementary products will be greater when users are farther away from a gate, allowing the player to make more progress before reaching the gate. Furthermore, because these users have a greater chance of experiencing progress (progress that would be prevented had they been closer to a gate), these users are also likely to increase their engagement in terms of # rounds and # playdays. We operationalize this variable using the number of levels until the next gate.

4.5.2 Impact of the intervention on customer targets

We evaluate the long-term impact of the intervention for these groups of customers by looking at the heterogeneity in treatment effect. We do so by running a set of linear regressions in which the dependent variable is the outcome of interest 30 days after the experiment, and as independent variables we include the treatment dummy, the player heterogeneity variable, and the interaction between these two. We also include a control variable for the amount of play during the 7 days before the start of the experiment, capturing the intensity of difficulty reduction that users would have faced on the first day of the intervention. We do so to ensure that the reported difference between groups are not due to possible correlation between the heterogeneity variable and treatment intensity (we standardize the variables to make the magnitude of the coefficients comparable). Finally, we compute, for each regression, the ratio between the interaction term and the treatment effect to get a measure of "how much the treatment effect increases for each type of customer" (Figure 7). The full set of regression results are presented in Appendix A6.

As expected, the users identified as potential targets (i.e., those whose heterogeneity variables are large and positive—one standard deviation above the mean), present stronger effects than the overall population. We find that *early_progress*, customers present effects that are approximately 26% stronger than those for the average customer on retention, engagement and monetization outcomes. The interaction is slightly weaker, yet, meaningful and significant, when progress is operationalized as *progress_prone*. The group of *spenders* show an interesting pattern. As predicted, these users do not show stronger effects in retention outcomes (e.g., similar levels of playing days, rounds and progress), but due to their preference for spending, they exhibit a very strong effect in monetization outcomes. Interestingly, *frustrated* users show the opposite pattern. While they respond more positively to the intervention when it comes to engagement (i.e., the impact on # rounds, # playdays and progress made), such an increase does not translate into stronger effects on monetization outcomes. Furthermore, *distance_to_ gate* not only moderates the effect of the intervention on engagement outcomes (27% increase in #playdays, 34% increase in # rounds), but also moderates, even more strongly, the effect on monetization (with 70% increase in #purchases, 59% increase in the user of coins). These results highlight the second effect of retention: When users are farther from a gate they can benefit more from progress and thus respond more positively in terms of engagement. However, they increase monetization further because the (additional) progress enjoyed (due to the intervention) encourages them to spend money to continue making progress.

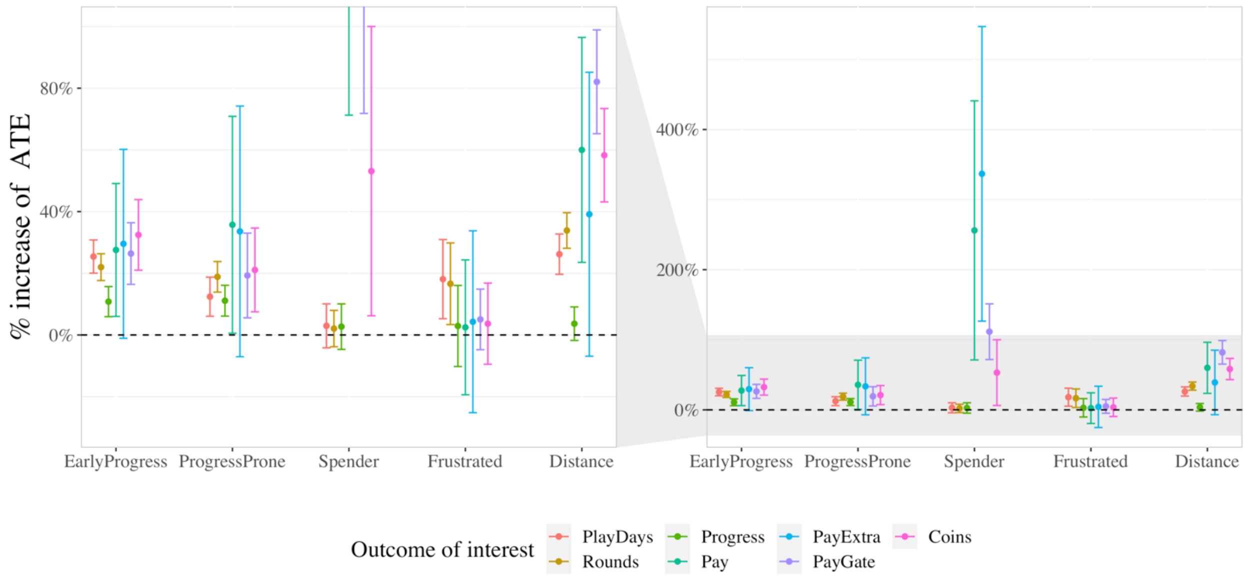


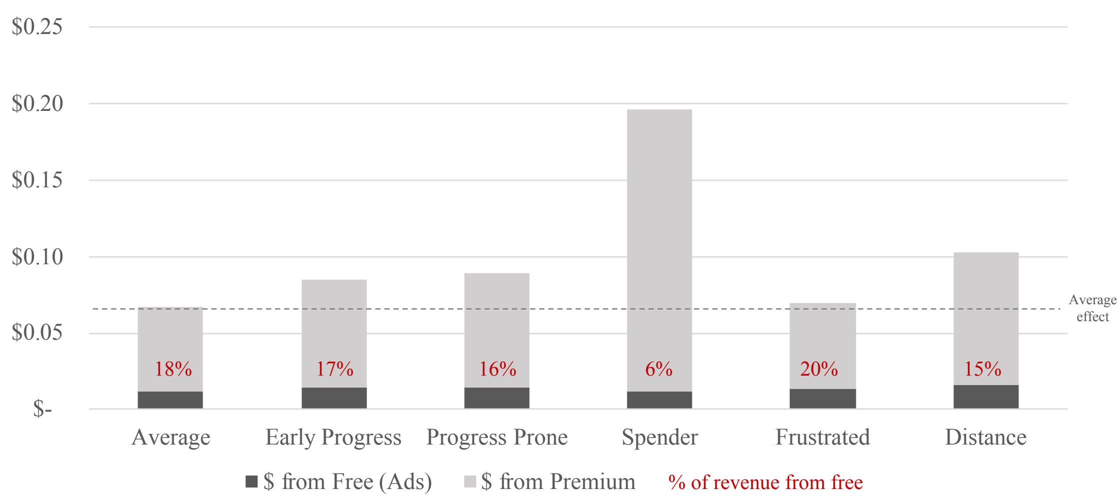
Figure 7: Increase in treatment effect for selected group of customers. *This figure shows the percentage increase in the effect of the intervention (on each outcome) for a standard deviation change in the heterogeneity variable. For example, the first (red) mark indicates that the intervention is 25% stronger (in amount of playdays in the next 30 days) when the "early_progress" is one standard deviation above the population mean. The left panel is a zoomed-in version of the right panel to better appreciate the differences.*

Overall, our targeting analyses help to further pinpoint the mechanism we propose of the effect of difficulty on game outcome through game progress. We demonstrate that groups of players who are theoretically most likely to be affected by the treatment due to need for progress or tendency to spend indeed demonstrate the strongest treatment effect. From a managerial point of view, we highlight that certain players (and at certain timing) show a much stronger effect (dozens and sometimes hundreds percent stronger) to reduce game difficulty, which highlights the opportunity in personalizing product design in freemium settings.

5 REVENUE DECOMPOSITION

We have demonstrated the effect of giving more of the free product (reduced difficulty) on engagement, retention, and monetization. One way to combine all of these effects together is to translate them to monetary values in terms of revenue for the firm. Thus, in this section, we quantify the net effect of the intervention (in monetary terms) and decompose it into the extra revenue obtained from the free (i.e., advertising) and from the premium (i.e., additional purchases) sides of the service, and compare the effect decomposition of the average player with that of the "target" users. Because we do not observe advertising revenue from the focal company and the exact monetization rates (from premium services) was disguised for confidentiality, we use average figures from the industry and reports from the focal company to quantify revenue for this analysis. In particular, we assume that the revenue for an ad exposure is 1.40¢ per round and that the average expenditure, given purchase is $4.21 per purchase.[25]

We combine the monetization estimates with the game play data and our estimates of treatment effect to compute the net and disaggregated effect of the treatment, in USD. Specifically, because we observe the rounds played and purchases made by each player during the 30 days following the experiment, we multiply the ad revenue per round with the lift in number of rounds, and the average purchase amount per purchase with the lift in number of purchases made due to the intervention. The results are presented Figure 8.



[25] We corroborate with members of the focal firm that these figures are representative of their business. The average expenditure per paying round is obtained from our data and the advertising revenue can be thought as an average of $7 per thousand exposures (CPM) and a single ad being shown every 5 rounds.

Figure 8: Decomposition in treatment effect for the average customer and for selected groups of customers. *This figure shows the additional revenue generated by the intervention. The numbers in red correspond to the proportion of additional revenue that comes from the advertising side. To be read: for an average customer, the intervention increased total revenue by $0.07. 18% of this revenue increase comes from advertising.*

Given our back-of-the-envelope calculations, the intervention generated an additional (net) revenue of $0.07 per customer in the 30 days after the intervention. Given the millions of customers who are eligible for the intervention at the firm at any given point in time, this increase is financially substantial. Of that revenue increase, 18% comes from the additional revenue that the company would collect from serving ads to customers during their extended lifetime while the other 82% comes from the additional purchases made by these customers. The fact that the majority of the treatment effect comes from in-app spending may be surprising given that the lever used in this retention effort gave players an easier game setting, which in theory should have cannibalized their in-app purchases because players often buy in-app extras in order to pass difficult levels. However, consistent with the reported twofold effect of retention in freemium settings, we observe that the majority of increase in revenue from the retention effort came from increased purchases.

Comparing these figures across "target" groups, we can further "monetize" the heterogeneity in treatment effect analysis. The group that offers the least benefit with respect to the average user is the frustrated type, whereas "spenders" generate more than twice the additional revenue. For those users, 94% of the extra revenue is coming from the purchase of premium items. While these "back-of-the-envelope" calculations are only approximative in that they are based on general industry revenue figures, we believe that this analysis highlights the importance of incorporating the impact of retention on short- and long-term monetization. Furthermore, it also allows us to combine together the different effects of changing the freemium product and better understand differences across customers.

6 DISCUSSION

We investigate the role of customer retention in a mobile game freemium context. Whereas much of the literature on the freemium business model has focused on the tradeoff between customer acquisition and monetization (via either paid services or advertising), we also bring into consideration the role of retention. We argue and demonstrate that in the context of IAPs freemium models, commonly used in the online games industry, enhancing the free portion of product (making the game easier) to customers at risk of churning, not only increases engagement and retention of players, but can also increase game monetization from in-app purchases because these retained customers have more opportunities to spend. We find that this dual effect of retention exists not only in the long-run but even in the short-term. Specifically, we use a field experiment in which game difficulty was exogenously manipulated; customers who did not play extensively in the past week were randomly allocated to a treatment condition with a reduced game difficulty, which can substitute buying extras to make the game easier. We find that the firm may have been overly myopic in setting game difficulty and not adjusting it to prevent customer churn. Reducing game difficulty among customers at risk leads not only to increased engagement and retention but also to an increase in in-app spending even within the first day of difficulty reduction. Financially, the net effect translated to an additional revenue of $0.07 per user, with around 82% of that extra revenue coming from the monetization of premium services.

Whereas we find that the effect of game difficulty is overall positive among customers at risk of churning, we also find substantial heterogeneity in the effect of reducing game difficulty, where customers who are more prone to making progress in the game exhibit stronger effects.

Moreover, surprisingly we find that customers who previously spent money on the game exhibit the strongest effect on in-app purchases. The firm may be reluctant to cannibalize spending by giving to these “spenders” customers more of the free product, but interestingly, for these users, giving them more services “for free” (i.e., making that game easier) increases their expenditure more than twice the increase of the average user. Furthermore, at times of frustration with progress and distance away from gates (when more progress can be made without a block) the effect of reducing game difficulty is even stronger. Using both heterogeneity in treatment effect and IV regressions of game progress on customer behavior (with treatment as an exogenous instrument), we show that the mechanism for the positive effect of reduced game difficulty on engagement and retention is through consumers who strive for game progress.

From a theoretical point of view, our research demonstrates the dual effect of customer retention in balancing the tradeoff between the free and the paid aspect of freemium products. Whereas our analyses focused on online gaming and game difficulty, we encourage future work to explore the role of retention in other freemium settings and levers other than game difficulty in striking the balance between the free and premium parts of the product or service. To the best of our knowledge this is the first paper to empirically explore the role of retention and retention effort in freemium settings.

From a substantive point of view, we demonstrate a substantial increase in engagement retention and IAPs due to adjustment of game difficulty. We believe that companies leave money on the table by not considering (or underestimating) the effect of retention in freemium settings. Furthermore, this work provides strong empirical support that personalizing the freemium product at the individual level and at the right timing can further increase financial profitability. Whereas in our empirical analysis we did not find individuals for whom reducing game difficulty would lead to a negative effect — possibly due to the focus on customers at the risk of churning — we expect that making the game too easy for individuals who seek challenges in the game may backfire. In such cases the heterogeneity in treatment effect discussed in Section 4.5 is even more important.

One of the commonly discussed factors in the context of freemium economy is network externality, in which the use of the free product by existing customers benefit other users as well (e.g., Shi, Zang and Srinivasan 2019). Because the game we work with is largely an individual play game, we do not consider the possible effect of network externality. However, to the extent that such network externalities exist, we expect the positive effect of increasing customer retention to be even stronger due to the social multiplier of increased engagement and retention of the targeted or treated users (Ascarza et al. 2017; Godinho de Matos, Pedro and Rodrigo 2018). Similar arguments can be made for the word of mouth effect of existing customers on the acquisition of future customers.

An interesting research direction building on this work is to examine whether the strategy of adjusting difficulty can be used as a long-term strategy for the firm. For example, do users' sensitivities to the adjusted difficulty change over time? And what happens to treated users if/when difficulty goes back to regular levels? Does their behavior resemble that of users who were never treated, or did the intervention prevent these users from developing skills in the game affecting their performance down the road? Beyond online gaming, this area of research would be relevant for contexts such as education and training settings where the level of difficulty will determine engagement and therefore the ability to learn.

In sum, our research demonstrates the dual effect of retention in freemium settings. More generally, we encourage researchers and practitioners in freemium settings to take a more holistic customer relationship management view which goes beyond acquisition and short-term gain when designing and evaluating freemium products and services.

REFERENCES

Amabile, T., & Kramer, S. (2011). The progress principle: Using small wins to ignite joy, engagement, and creativity at work. Harvard Business Press.

Appel, G., Libai, B., Muller, E., & Shachar, R. (2020). Retention and the Monetization of Apps. SSRN

Arora, N., Dreze, X., Ghose, A., Hess, J.D., Iyengar, R., Jing, B., Joshi, Y., Kumar, V., Lurie, N., Neslin, S. and Sajeesh, S., (2008). Putting one-to-one marketing to work: Personalization, customization, and choice. Marketing Letters, 19(3-4), p.305.

Ascarza, E., Ebbes, P., Netzer, O. and Danielson, M., (2017). Beyond the target customer: Social effects of customer relationship management campaigns. Journal of Marketing Research, 54(3), 347-363.

Ascarza, E., Iyengar, R., & Schleicher, M. (2016). The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. Journal of Marketing Research, 53(1), 46-60.

Ascarza, E., Neslin, S.A., Netzer, O., Anderson, Z., Fader, P.S., Gupta, S., Hardie, B.G.S., Lemmens, A., Libai, B., Neal, D., Provost, F., & Schrift, R. (2018), In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions. Customer Needs and Solutions 5:65–81.

Bapna, R., Ramaprasad, J. & Umyarov, A. (2016) Monetizing Freemium Communities: Does Paying for Premium Increase Social Engagement? Available at http://dx.doi.org/10.2139/ssrn.2885681.

Bapna R, Umyarov A (2015) Do your online friends make you pay? a randomized _eld experiment on peer influence in online social networks. *Management Science* 61(8):1902-1920.

Cheng, H. K., & Tang, Q. C. (2010). Free trial or no free trial: Optimal software product design with network effects. *European Journal of Operational Research*, 205(2), 437–447.

Chiou, L., & Tucker, C. (2013). Paywalls and the demand for news. Information Economics and Policy, 25(2), 61–69.

Deng, Y., Lambrecht, A., & Liu, Y. (2020). Spillover effects and freemium strategy in mobile app market. Available at SSRN 3149550.

Forbes (2015), "Should You Consider A Freemium Model For Your Business?," Available at: https://www.forbes.com/sites/chuckcohn/2015/07/02/should-you-consider-a-freemium-model-for-your-business-pros-cons/#7cd2876624a7. Last accessed 2020-03-16.

Godinho de Matos, M., Pedro, F. and Rodrigo, B., 2018. Target the ego or target the group: Evidence from a randomized experiment in proactive churn management. Marketing Science, 37(5), pp.793-811.

Gu, X., Kannan, P. K., & Ma, L. (2018). Selling the premium in freemium. Journal of Marketing, 82(6), 10-27.

Gupta, S. and Zeithaml, V., 2006. Customer metrics and their impact on financial performance. Marketing science, 25(6), pp.718-739.

Hadiji, F., Sifa, R., Drachen, A., Thurau, C., Kersting, K., & Bauckhage, C. (2014, August). Predicting player churn in the wild. In 2014 IEEE Conference on Computational Intelligence and Games (pp. 1-8). Ieee.

Halbheer, D., Stahl, F., Koenigsberg, O., & Lehmann, D. R. (2014). Choosing a digital content strategy: How much should be free?. International Journal of Research in Marketing, 31(2), 192-206.

Hofacker, C. F., De Ruyter, K., Lurie, N. H., Manchanda, P., & Donaldson, J. (2016). Gamification and mobile marketing effectiveness. Journal of Interactive Marketing, 34, 25-36.

Huang, Y., Jasin, S., & Manchanda, P. (2019). "Level Up": Leveraging Skill and Engagement to Maximize Player Game-Play in Online Video Games. Information Systems Research, 30(3), 927-947.

Kumar, V. (2014). Making" freemium" work. Harvard business review, 92(5), 27-29.

Lambrecht, A., & Misra, K. (2017). Fee or free: when should firms charge for online content?. *Management Science*, 63(4), 1150-1165.

Lee, S. K., Hong, S. J., Yang, S. I., & Lee, H. (2016, October). Predicting churn in mobile free-to-play games. In 2016 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 1046-1048). IEEE.

Lee, C., Kumar, V., & Gupta, S. (2017). Designing Freemium: Strategic Balancing of Growth and Monetization. Working paper.

Lemmens, A., & Gupta, S. (2020). Managing churn to maximize profits. *Marketing Science*, Forthcoming.

Levitt, S. D., List, J. A., Neckermann, S., & Nelson, D. (2016). Quantity discounts on a virtual good: The results of a massive pricing experiment at King Digital Entertainment. *Proceedings of the National Academy of Sciences*, *113*(27), 7323-7328.

Li, H., Jain, S., & Kannan, P. K. (2019). Optimal Design of Free Samples for Digital Products and Services. *Journal of Marketing Research*, *56*(3), 419-438.

Liu, C. Z., Au, Y. A., & Choi, H. S. (2014). Effects of freemium strategy in the mobile app market: An empirical study of google play. *Journal of Management Information Systems*, *31*(3), 326-354.

Oestreicher-Singer G, Zalmanson L (2013) Content or community? a digital business strategy for content providers in the social age. Management Information Systems Quarterly 37(2), 591-661.

Padilla, N., Ascarza, E. (2020) Overcoming the Cold Start Problem of CRM using a Probabilistic Machine Learning Approach. Available at SRN: https://ssrn.com/abstract=2933291

Runge, J., Gao, P., Garcin, F., & Faltings, B. (2014, August). Churn prediction for high-value players in casual social games. In *2014 IEEE conference on Computational Intelligence and Games* (pp. 1-8). IEEE.

Runge, J., Nair, H., & Levav, J. (2019). Price Promotions in "Freemium" Settings. *Stanford GSB Working Paper No. 3769.*

Runge, J., Wagner, S., Claussen, J., & Klapper, D. (2016). Freemium pricing: Evidence from a large-scale field experiment. *Humboldt University Berlin, School of Business and Economics, Institute of Marketing Working Paper*.

Seufert, E. B. (2013). *Freemium economics: Leveraging analytics and user segmentation to drive revenue*. Elsevier.

Shi, Z., Zhang, K., & Srinivasan, K. (2019). Freemium as an Optimal Strategy for Market Dominant Firms. *Marketing Science*, *38*(1), 150-169.

Sifa, R., Hadiji, F., Runge, J., Drachen, A., Kersting, K., & Bauckhage, C. (2015). Predicting purchase decisions in mobile free-to-play games. In Eleventh Artificial Intelligence and Interactive Digital Entertainment Conference.

Sifa, R., Runge, J., Bauckhage, C., & Klapper, D. (2018). Customer Lifetime Value Prediction in Non-Contractual Freemium Settings: Chasing High-Value Users Using Deep Neural Networks and SMOTE. Proceedings of the 51st Hawaii International Conference on System Science.

Sun, T., Gao, G., & Jin, G. Z. (2019). Mobile Messaging for Offline Group Formation in Prosocial Activities: A Large Field Experiment. *Management Science*, *65*(6), 2717-2736.

Sun, T., Viswanathan, S., & Zheleva, E. (2020). Creating social contagion through firm-mediated message design: Evidence from a randomized field experiment. Forthcoming at *Management Science*.

Viljanen, M. J., Airola, A., Majanoja, A. M., Heikkonen, J., & Pahikkala, T. (2020). Measuring player retention and monetization using the mean cumulative function. IEEE Transactions on Games.

Voigt, S., & Hinz, O. (2016). Making digital freemium business models a success: Predicting customers' lifetime value via initial purchase information. Business & Information Systems Engineering, 58(2), 107-118.

Xie, H., Devlin, S., Kudenko, D., & Cowling, P. (2015, August). Predicting player disengagement and first purchase with event-frequency based data representation. In 2015 IEEE Conference on Computational Intelligence and Games (CIG) (pp. 230-237). IEEE.

APPENDIX

A1 Experimental design – Further details about the degree of difficulty

Table A 1 shows the treatment intensity for each group of players, depending on the number of rounds that they have played in the previous 7 days. If a user played 20 or more rounds, the difficulty remained intact, at the default level denoted by 5. If a user played between 15 and 19 rounds, the difficultly was set to 4, which implies that the probability of connectivity in the board increased by 10% (i.e., the game was made easier for that player). As the user played fewer rounds in the last 7 days, the chances of connecting pieces was made easier.

# rounds in last 7 days	Difficulty level	Increase of "connectivity"
>= 20	5 (more difficult)	0%
< 20	4	10%
< 15	3	20%
< 10	2	30%
< 5	1 (least difficult)	45%

Table A 1: *Treatment intensity by level of play during the past 7 days. Difficulty 5 is the default option in the game design.*

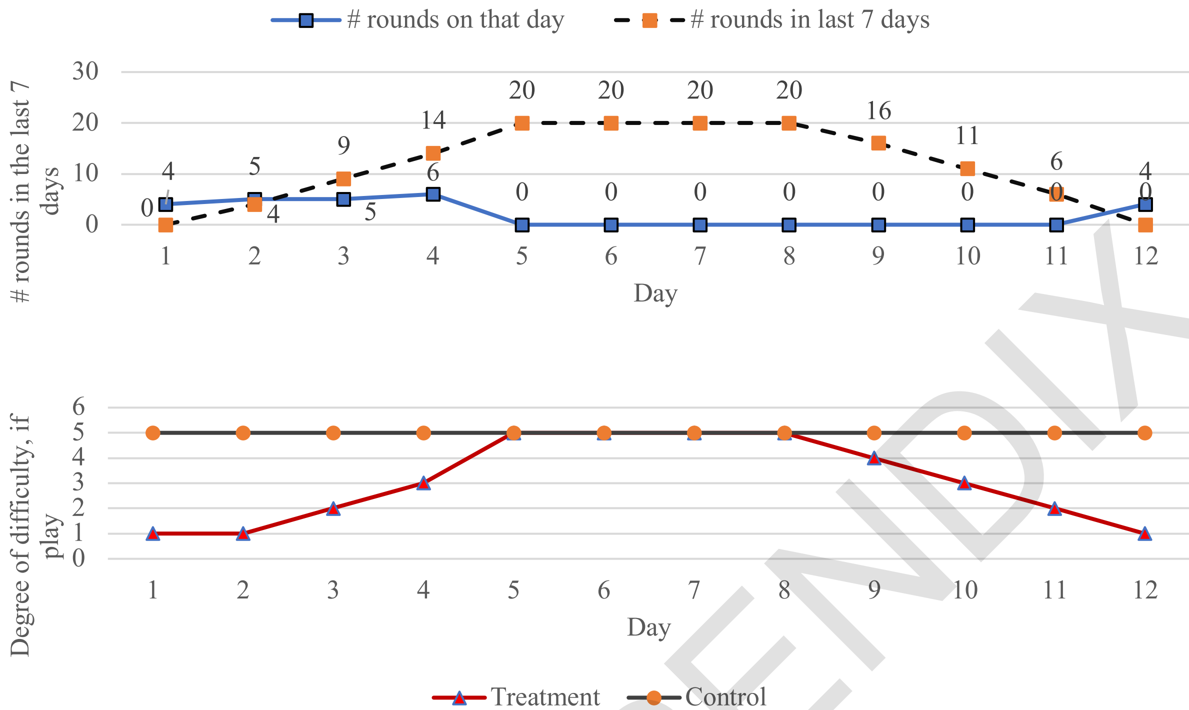


Figure **A1** shows a hypothetical example for the relationship between past play (top figure) and the degree of difficulty (bottom figure). The red dashed line corresponds to users in the treatment condition, whose difficulty is manipulated depending on the number of rounds played in the last 7 days, the black-solid line corresponds to users in the control condition, whose difficulty is set always to the highest (default) level.

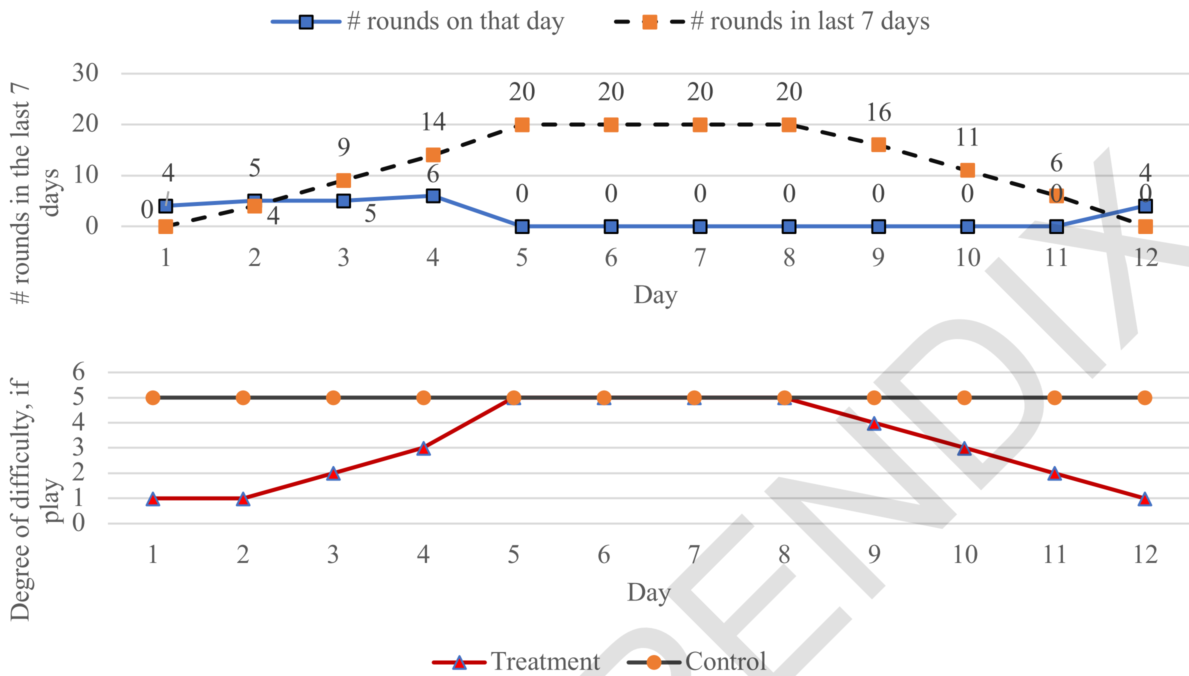


Figure A1: *Hypothetical example to illustrate the level of difficulty (bottom figure) that correspond to different levels of past play. The red line (dashed) correspond to treated users, whose difficulty was altered based on the amount of play in the previous 7 days, and the back line (solid) corresponds to control users whose level of difficulty never changed.*

Figure A2 shows two difficulty scenarios. On the left the player is facing the highest level of difficulty ("default" in the game design) whereas the player on the right is facing the easiest scenario with low difficulty levels.

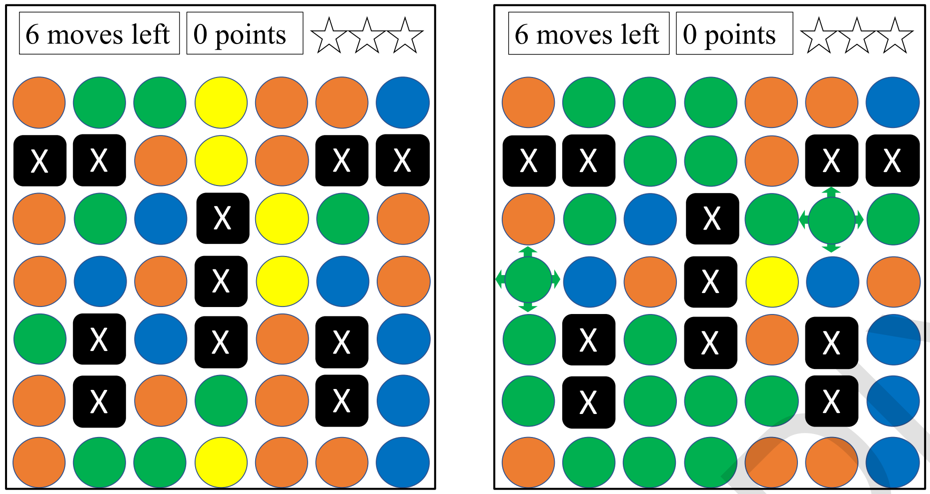


Figure A2: Example of difficulty manipulation. *Image of a level of the game, with two different levels of difficulty. On the left, the most difficult scenario which corresponds with the default difficulty of the game. On the right, the player is facing an easier scenario in which there are much higher chances of creating long combinations (of green color) and has two "special chips," identified by the arrows around the two pieces.*

A2 Data – Descriptive statistics and randomization checks

Table A **2** shows the descriptive statistics and randomization checks for the full set of user-level variables. Level20 variables are measured when users pass level 20 (and remains constant over time) and all other variables are measured at the moment of the intervention.

N=	329,999					190,863	139,136		
	Mean	SD	p25	p50	p75	Control	Treatment	Difference	p-value
level20_rounds	45.83	36.91	29	36	49	45.85	45.81	-0.04	0.810
level20_days	8.691	15.74	1	3	9	8.67	8.719	0.049	0.373
level20_stars	39.80	4.497	37	39	42	39.81	39.78	-0.03	0.110
level20_coins_collected	40.75	75.93	0	0	70	40.82	40.64	-0.18	0.499
level20_did_use_extra	0.926	0.261	1	1	1	0.927	0.926	-0.001	0.668
level20_did_use_coin	0.513	0.5	0	1	1	0.513	0.514	0.001	0.343
level20_did_purchase	0.011	0.103	0	0	0	0.011	0.011	0.000	0.080
age_rounds	189.6	217.7	62	113	225	190.0	189.2	-0.8	0.315
age_days	45.85	31.91	19	37	66	45.78	45.94	0.16	0.168
age_level	37.74	16.97	24	39	40	37.78	37.69	-0.09	0.101
age_stars	75.19	36.68	49	67	87	75.3	75.04	-0.26	**0.045**
age_coins	24.0	74.0	0	0	30	24.04	23.96	-0.08	0.752
age_distance_to_gate	6.353	6.561	0	4	11	6.484	6.174	-0.31	**0.000**
rfm_rec	13.88	17.9	3	7	16	13.85	13.92	0.07	0.276
rfm_week	5.728	6.68	0	2	12	5.735	5.718	-0.017	0.476
rfm_ratio	3.719	2.952	1.667	3.077	4.836	3.721	3.716	-0.005	0.605
stuck_rounds	26.72	62.45	2	7	25	26.79	26.61	-0.18	0.403
stuck_days	17.25	21.59	4	9	21	17.22	17.3	0.08	0.347
stuck_playdays	3.054	3.804	1	2	4	3.054	3.055	0.001	0.930
stuck_days_in_gate	23.47	23.38	8	15	32	23.32	23.67	0.35	**0.033**
stuck_broke_gate	0.0218	0.146	0	0	0	0.0217	0.022	0.0003	0.645
yesterday_progress	1.225	2.564	0	0	2	1.223	1.227	0.004	0.703
yesterday_win_prop	0.308	0.341	0	0.2	0.545	0.309	0.307	-0.002	0.289
skill_rounds_per_level	4.41	3.653	2.227	3.286	5.318	4.413	4.406	-0.007	0.568
skill_stars_per_level	1.988	0.257	1.821	1.957	2.149	1.988	1.987	-0.001	**0.036**

Table A 2: *Descriptive statistics (left-most columns) and randomization checks (last 4 columns).*

A3 Manipulation checks by treatment intensity

Table A3 shows the remaining metrics examined in the manipulation checks.

	Points (by level of difficulty)				Snake Length (by level of difficulty)			
	4	3	2	1	4	3	2	1
Treatment	**1275**	**3225**	**6126**	**10741**	0.018*	**0.027**	**0.112**	**0.269**
	(128.1)	(137.9)	(157.4)	(91.2)	(0.009)	(0.009)	(0.009)	(0.005)
Constant	35818	35862	35648	35047	5.119	5.148	5.144	5.179
	(82.7)	(87.6)	(95.8)	(50.8)	(0.006)	(0.006)	(0.007)	(0.003)
N	370137	302836	281827	1055166	370137	302836	281827	1055166
R-squared	0.1%	0.6%	2.0%	4.9%	0.0%	0.0%	0.2%	1.1%

Table A3: *Manipulation checks by degree of difficulty. OLS of the round outcome against a treatment dummy using all rounds on the first day of the experiment. Standard errors are clustered at the user level. Treatment variable in bold indicates that p-value<0.01, * indicates that p-value<0.05.*

A4 Robustness checks for the main analyses

In Table A4 we replicate the results presented in Tables 6 through 8 of the main manuscript only using the first 5 observations per user. With the exception of coin extra and money extra (for which the estimate for the treatment effect becomes insignificant), all other results are consistent with those obtained in the main analysis.

	Continue playing	Use extras	Use coins	Use Money	Coin extra	Coin gate	Money extra	Money gate
Treatment	**0.0145**	**0.0012**	**0.0004**	**0.0002**	-0.00026	**0.00235**	0.00002	**0.00066**
	(0.0008)	(0.0004)	(0.0002)	(0.0001)	(0.00014)	(0.00036)	(0.00006)	(0.00016)
p-val	0.0000	0.0016	0.0269	0.0155	0.06340	0.00000	0.74800	0.00004
Constant	0.7920	0.0261	0.0095	0.0015	0.00503	0.00875	0.00078	0.00163
	(0.0005)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0)	(0.0001)
# obs	1136686	1177102	1177102	1177102	1177102	315600	1177102	315600

Table A4: Robustness check using first 5 rounds per user. *OLS of the behavior of interest against a treatment dummy using the first five observations per user. Standard errors are clustered at the user level. The gate-related outcomes are conditioned on the user being at a gate Treatment variable in bold indicates that p-value<0.05.*

Similarly, Table A5 shows the results when we add multiple controls per user. We use as controls all the observables at the moment in which the user receives the experiment for the first time. All results are consistent with those presented in Tables 6 through 8 of the main analyses.

	Continue playing	Use extras	Use coins	Use Money	Coin extra	Coin gate	Money extra	Money gate
Treatment	**0.0202**	**-0.0016**	**-0.0006**	-0.0001	**-0.00062**	-0.00010	**-0.00012**	0.00001
	(0.0007)	(0.0003)	(0.0001)	(0.0001)	(0.00011)	(0.00019)	(0.00005)	(0.00008)
p-val	*0.0000*	*0.0000*	*0.0001*	*0.0660*	*0.00000*	*0.58000*	*0.01040*	*0.93200*
Constant	0.7850	0.0033	0.0021	-0.0109	-0.00821	0.03160	-0.00889	-0.00976
	(0.0083)	(0.004)	(0.003)	(0.0013)	(0.0021)	(0.00478)	(0.00103)	(0.00347)
# obs	1867811	2009921	2009921	2009921	2009921	652565	2009921	652565

Table A5: Robustness check adding controls. *OLS of the behavior of interest against a treatment dummy using all rounds on the first days of the intervention. Standard errors are clustered at the user level. The gate-related outcomes are conditioned on the user being at a gate Treatment variable in bold indicates that p-value<0.01.*

A5 Robustness checks for the IV regressions

We replicate the analyses presented in Section 4.4 using different proxies for progress.

Specifically, we compute:

- "progress so far today" which corresponds to the number of new levels a user had completed successfully up to that point that day (Table A6),
- "progress last round" which captures whether the user made progress in the most recent round (Table A7),
- "progress today vs. the past" which is the ratio between the progress made so far today and the average daily progress made by the user prior to the intervention (Table A8), and
- "proportion of progress today vs. the past" which is the ratio between the proportion of new levels in which the user made progress so far and the average of such a metric before the intervention – i.e., it is a ratio of proportions (Table A9).

All results are consistent with those presented in the main manuscript, confirming that progress is an important driver in the effects we find.

	Next day retention	Continue playing	Use extras	Coin gate	Coin extra	Money gate	Money extra
Progress so far today	**0.0314**	**0.0126**	**-0.0030**	**0.0007**	**-0.0007**	**0.0001**	**-0.0001**
	(0.002)	(0.0006)	(0.0003)	(0.0001)	(0.0001)	(0)	(0)
p-val	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0429
Constant	0.2920	0.8400	0.0260	0.0003	0.0049	0.0000	0.0008
	(0.0032)	(0.0009)	(0.0005)	(0.0001)	(0.0002)	(0)	(0.0001)
# obs	261534	1802322	1679967	1679967	1679967	1679967	1679967

***Table A6: Robustness check (I) for the impact of progress on behavioral outcomes**. 2SLS regression of the behavior of interest against a variable that captures whether the user made progress on the most recent round. "Progress so far today" is instrumented with the treatment variable. Standard errors are clustered at the user level. The number of observations corresponds to the rounds used for each regression. Treatment variable in bold indicates that p-value<0.05.*

	Next day retention	Continue playing	Use extras	Coin gate	Coin extra	Money gate	Money extra
Progress last round	**0.3980**	**0.2490**	**-0.0367**	**0.0082**	**-0.0081**	**0.0014**	**-0.0012**
	(0.0249)	(0.0095)	(0.0038)	(0.0007)	(0.0016)	(0.0003)	(0.0006)
p-val	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0429
Constant	0.2620	0.7800	0.0288	-0.0003	0.0055	-0.0001	0.0009
	(0.0029)	(0.0019)	(0.0008)	(0.0001)	(0.0003)	(0.0001)	(0.0001)
# obs	326472	1867849	1679967	1679967	1679967	1679967	1679967

***Table A7: Robustness check (II) for the impact of progress on behavioral outcomes**. 2SLS regression of the behavior of interest against a variable that captures whether the user made progress on the most recent round. "Progress last round" is instrumented with the treatment variable. Standard errors are clustered at the user level. The number of observations corresponds to the rounds used for each regression. Treatment variable in bold indicates that p-value<0.05.*

	Next day retention	Continue playing	Use extras	Coin gate	Coin extra	Money gate	Money extra
Progress today vs. past	**0.0920**	**0.0410**	**-0.0097**	**0.0022**	**-0.0021**	**0.0004**	**-0.0003**
	(0.0059)	(0.0021)	(0.001)	(0.0002)	(0.0004)	(0.0001)	(0.0002)
p-val	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0428
Constant	0.2950	0.8410	0.0256	0.0004	0.0048	0.0001	0.0007
	(0.003)	(0.0009)	(0.0004)	(0.0001)	(0.0002)	(0)	(0.0001)
# obs	261534	1802322	1679967	1679967	1679967	1679967	1679967

***Table A 8: Robustness check (III) for the impact of progress on behavioral outcomes**. 2SLS regression of the behavior of interest against a variable that captures whether the user made progress on the most recent round. "Progress today vs. past" is instrumented with the treatment variable. Standard errors are clustered at the user level. The number of observations corresponds to the rounds used for each regression. Treatment variable in bold indicates that p-value<0.05.*

	Next day retention	Continue playing	Use extras	Coin gate	Coin extra	Money gate	Money extra
Prop. progress today vs. past	**0.0499**	**0.0195**	**-0.0071**	**0.0014**	**-0.0016**	**0.0003**	**-0.0002**
	(0.0035)	(0.001)	(0.0006)	(0.0001)	(0.0002)	(0)	(0.0001)
p-val	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0288
Constant	0.3070	0.8290	0.0374	-0.0002	0.0070	-0.0001	0.0011
	(0.0035)	(0.0013)	(0.0007)	(0.0001)	(0.0003)	(0)	(0.0001)
# obs	188783	1191700	1116023	1116023	1116023	1116023	1116023

***Table A 9: Robustness check (IV) for the impact of progress on behavioral outcomes**. 2SLS regression of the behavior of interest against a variable that captures whether the user made progress on the most recent round. "Prop. progress today vs. past" is instrumented with the treatment variable. Standard errors are clustered at the user level. The number of observations corresponds to the rounds used for each regression. Treatment variable in bold indicates that p-value<0.05.*

A6 Implications for freemium design: Full set of results

In this appendix we report the full set of results from the regression analyses conducted in Section 5 when exploring the groups of customers who would be more responsive to the intervention.

	Playdays	Rounds	Progress	Purchases	Pay extra	Pay gate	Coins	Coin extra	Coin gate
Treatment	0.5900	8.4800	2.3210	0.0131	0.0052	0.0053	0.0293	0.0112	0.0187
	(0.0192)	(0.234)	(0.0623)	(0.0024)	(0.0012)	(0.0004)	(0.0022)	(0.0018)	(0.0008)
p-val	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Interaction	**0.1500**	**1.8650**	**0.2510**	**0.0036**	**0.0015**	**0.0014**	**0.0095**	**0.0042**	**0.0046**
	(0.0162)	(0.188)	(0.0577)	(0.0014)	(0.0008)	(0.0003)	(0.0017)	(0.0014)	(0.0007)
p-val	*0.0000*	*0.0000*	*0.0000*	*0.0121*	*0.0585*	*0.0000*	*0.0000*	*0.0030*	*0.0000*
Main Effect	0.2710	3.1360	4.9340	0.0114	0.0061	0.0016	0.0178	0.0111	0.0051
	(0.0104)	(0.119)	(0.0363)	(0.0009)	(0.0005)	(0.0002)	(0.0011)	(0.0009)	(0.0004)
p-val	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Constant	1.6470	13.3300	2.5480	0.0197	0.0120	0.0027	0.0363	0.0271	0.0086
	0.0103	0.1250	0.0322	0.0012	0.0006	0.0002	0.0012	0.0010	0.0004
Intensity Group	5.3650	38.8000	41.6700	0.0476	0.0247	0.0080	0.1560	0.0960	0.0385
	(0.012)	(0.145)	(0.0389)	(0.0014)	(0.0007)	(0.0002)	(0.0013)	(0.0011)	(0.0005)
# obs	329999	329999	329999	329999	329999	329999	329999	329999	329999
Ratio: Interaction / Treatment	25.4%	22.0%	10.8%	27.6%	29.6%	26.4%	32.5%	37.6%	24.6%

***Table A 10: Interaction effect with Early_Progress**. OLS with interaction effect. Variable has been standardized. Robust standard errors reported in parentheses. Interaction variable in bold indicates that p-value<0.06.*

	Playdays	Rounds	Progress	Purchases	Pay extra	Pay gate	Coins	Coin extra	Coin gate
Treatment	0.5900	8.4960	2.3220	0.0131	0.0052	0.0053	0.0293	0.0112	0.0187
	(0.0192)	(0.234)	(0.0638)	(0.0024)	(0.0012)	(0.0004)	(0.0022)	(0.0018)	(0.0008)
p-val	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Interaction	**0.0732**	**1.6020**	**0.2580**	**0.0047**	0.0018	**0.0010**	**0.0062**	**0.0038**	**0.0023**
	(0.0191)	(0.215)	(0.0592)	(0.0024)	(0.0011)	(0.0004)	(0.002)	(0.0017)	(0.0008)
p-val	*0.0001*	*0.0000*	*0.0000*	*0.0463*	*0.1050*	*0.0062*	*0.0023*	*0.0251*	*0.0028*
Main Effect	0.1510	4.2780	3.0540	0.0075	0.0039	0.0012	0.0069	0.0056	0.0024
	(0.012)	(0.133)	(0.0371)	(0.0018)	(0.0007)	(0.0002)	(0.0013)	(0.0011)	(0.0005)
p-val	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Constant	1.6380	13.2300	2.4270	0.0194	0.0118	0.0027	0.0358	0.0268	0.0084
	0.0103	0.1240	0.0329	0.0012	0.0006	0.0002	0.0012	0.0010	0.0004
Intensity Group	5.3650	38.7900	41.6700	0.0476	0.0247	0.0080	0.1560	0.0960	0.0385
	(0.012)	(0.145)	(0.0399)	(0.0014)	(0.0007)	(0.0002)	(0.0013)	(0.0011)	(0.0005)
# obs	329999	329999	329999	329999	329999	329999	329999	329999	329999
Ratio: Interaction / Treatment	12.4%	18.9%	11.1%	35.7%	33.6%	19.3%	21.1%	33.5%	12.5%

***Table A 11: Interaction effect with Progress_Prone**. OLS with interaction effect. Variable has been standardized. Robust standard errors reported in parentheses. Interaction variable in bold indicates that p-value<0.05.*

	Playdays	Rounds	Progress	Purchases	Pay extra	Pay gate	Coins	Coin extra	Coin gate
Treatment	0.5880	8.4630	2.3010	0.0123	0.0048	0.0053	0.0288	0.0107	0.0187
	(0.0192)	(0.235)	(0.0648)	(0.0023)	(0.0011)	(0.0004)	(0.0021)	(0.0018)	(0.0008)
p-val	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Interaction	0.0174	0.1760	0.0622	**0.0315**	**0.0163**	**0.0059**	**0.0153**	**0.0151**	**0.0029**
	(0.0214)	(0.255)	(0.0866)	(0.0116)	(0.0052)	(0.0011)	(0.0069)	(0.0063)	(0.0013)
p-val	*0.4160*	*0.4900*	*0.4720*	*0.0069*	*0.0017*	*0.0000*	*0.0268*	*0.0173*	*0.0271*
Main Effect	0.0826	0.3700	0.4670	0.0922	0.0458	0.0054	0.0658	0.0580	0.0047
	(0.0136)	(0.158)	(0.0546)	(0.0064)	(0.0029)	(0.0006)	(0.0041)	(0.0037)	(0.0008)
p-val	*0.0000*	*0.0190*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Constant	1.6380	13.2300	2.4240	0.0184	0.0113	0.0026	0.0352	0.0263	0.0084
	0.0103	0.1250	0.0334	0.0012	0.0006	0.0002	0.0012	0.0010	0.0004
Intensity Group	5.3650	38.8000	41.6800	0.0479	0.0248	0.0080	0.1560	0.0962	0.0385
	(0.012)	(0.145)	(0.0405)	(0.0014)	(0.0007)	(0.0002)	(0.0013)	(0.0011)	(0.0005)
# obs	329999	329999	329999	329999	329999	329999	329999	329999	329999
Ratio: Interaction / Treatment	3.0%	2.1%	2.7%	256.1%	336.8%	111.6%	53.1%	141.1%	15.5%

Table A 12: Interaction effect with Spender. *OLS with interaction effect. Variable has been standardized. Robust standard errors reported in parentheses. Interaction variable in bold indicates that p-value<0.03.*

	Playdays	Rounds	Progress	Purchases	Pay extra	Pay gate	Coins	Coin extra	Coin gate
Treatment	0.5910	8.4920	2.3120	0.0130	0.0052	0.0053	0.0292	0.0111	0.0187
	(0.0191)	(0.233)	(0.0642)	(0.0024)	(0.0012)	(0.0004)	(0.0022)	(0.0018)	(0.0008)
p-val	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Interaction	**0.1070**	**1.4100**	0.0675	0.0003	0.0002	0.0003	0.0011	0.0028	**-0.0023**
	(0.0387)	(0.573)	(0.155)	(0.0015)	(0.0008)	(0.0003)	(0.002)	(0.0016)	(0.0007)
p-val	*0.0058*	*0.0138*	*0.6630*	*0.8240*	*0.7750*	*0.3120*	*0.5860*	*0.0821*	*0.0006*
Main Effect	0.5620	8.1250	2.5490	-0.0051	-0.0030	-0.0007	-0.0111	-0.0090	-0.0018
	(0.0275)	(0.399)	(0.12)	(0.0007)	(0.0004)	(0.0001)	(0.0011)	(0.0009)	(0.0004)
p-val	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Constant	1.6120	12.8500	2.3160	0.0196	0.0119	0.0027	0.0363	0.0272	0.0086
	0.0103	0.1240	0.0332	0.0012	0.0007	0.0002	0.0012	0.0010	0.0004
Intensity Group	5.3640	38.7900	41.6800	0.0476	0.0247	0.0080	0.1560	0.0960	0.0385
	(0.0119)	(0.144)	(0.04)	(0.0014)	(0.0007)	(0.0002)	(0.0013)	(0.0011)	(0.0005)
# obs	329999	329999	329999	329999	329999	329999	329999	329999	329999
Ratio: Interaction / Treatment	18.1%	16.6%	2.9%	2.5%	4.3%	5.0%	3.7%	24.8%	-12.2%

***Table A 13: Interaction effect with Frustrated**. OLS with interaction effect. Variable has been standardized. Robust standard errors reported in parentheses. Interaction variable in bold indicates that p-value<0.02.*

	Playdays	Rounds	Progress	Purchases	Pay extra	Pay gate	Coins	Coin extra	Coin gate
Treatment	0.5880	8.4700	2.2710	0.0129	0.0051	0.0053	0.0290	0.0110	0.0186
	(0.0192)	(0.235)	(0.0634)	(0.0024)	(0.0012)	(0.0004)	(0.0021)	(0.0018)	(0.0008)
p-val	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Interaction	**0.1540**	**2.8690**	0.0831	**0.0077**	0.0020	**0.0044**	**0.0169**	0.0034	**0.0135**
	(0.0196)	(0.249)	(0.063)	(0.0024)	(0.0012)	(0.0005)	(0.0022)	(0.0018)	(0.0009)
p-val	*0.0000*	*0.0000*	*0.1870*	*0.0012*	*0.0947*	*0.0000*	*0.0000*	*0.0635*	*0.0000*
Main Effect	-0.0010	-2.1050	3.7320	0.0137	0.0079	0.0014	0.0151	0.0106	0.0052
	(0.012)	(0.149)	(0.0392)	(0.0013)	(0.0007)	(0.0002)	(0.0014)	(0.0011)	(0.0005)
p-val	*0.9340*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Constant	1.6360	13.2700	2.3030	0.0188	0.0115	0.0026	0.0351	0.0264	0.0081
	0.0103	0.1250	0.0326	0.0012	0.0006	0.0002	0.0012	0.0010	0.0004
Intensity Group	5.3650	38.7900	41.6900	0.0477	0.0247	0.0080	0.1560	0.0961	0.0385
	(0.012)	(0.145)	(0.0396)	(0.0014)	(0.0007)	(0.0002)	(0.0013)	(0.0011)	(0.0005)
# obs	329999	329999	329999	329999	329999	329999	329999	329999	329999
Ratio: Interaction / Treatment	26.2%	33.9%	3.7%	60.0%	39.1%	82.1%	58.3%	31.0%	72.6%

***Table A 14: Interaction effect with Distance**. OLS with interaction effect. Variable has been standardized. Robust standard errors reported in parentheses. Interaction variable in bold indicates that p-value<0.01.*