Consumer Product Discovery Costs, Entry, Quality and Congestion in Online Markets

Daniel Ershov*
Toulouse School of Economics
Université Toulouse 1 Capitole
daniel.ershov@tse-fr.eu

October 8, 2021

Abstract

This paper examines the effects of changing consumer product discovery costs in online markets on entry, market congestion, quality and consumer welfare. Using new Android app store data, I take advantage of a natural experiment that reduces discovery costs for game apps. I show that entry increases when discovery costs fall, potentially increasing congestion. Entrant quality falls. I disentangle the welfare effects of changing discovery costs and product assortment using a structural model that accounts for congestion externalities. I find that consumers' welfare increases but fifty percent of variety welfare gains are lost from higher congestion and lower entrant quality.

*This paper was a chapter in my PhD dissertation and previously circulated under the title "The Effects of Consumer Search Costs on Entry and Quality in the Mobile App Market." I would like to thank Victor Aguirregabiria, Avi Goldfarb and Heski Bar-Isaac for their time spent reading and discussing this paper. This paper substantially benefitted from discussions with Joel Waldfogel, Eduardo Souza Rodrigues, Mar Reguant, Sandro Shelegia and comments from seminar participants at many seminars and conferences including Toronto, TSE, Tilburg, Sciences Po, KU Leuven, Mannheim, CEMFI, EARIE and JEI, IIOC, the 8th Annual Workshop on Consumer Search and Switching Costs, the CSIO-IDEI workshop, the Barcelona GSE Summer Forum, CESifo and IO^2. I gratefully acknowledge past support received from SSHRC, the Ontario Government, the University of Toronto and ANR under grant ANR-17-EUR-0010 (Investissements d'Avenir program). All errors are my own.

1 Introduction

Although online marketplaces contain an enormous variety of products, discovering these is often challenging in practice. Search algorithms and other features of online marketplaces such as product categories, best-seller lists and ratings facilitate consumer discovery. Changes in these features could affect both demand-side consumer decisions and supply-side firm entry, product design and pricing decisions.[1] Economic theory predicts that a better discovery process should increase entry (Anderson and Renault 1999) and have ambiguous effects on product quality. But, paradoxically, increasing entry in response to easier product discovery can create congestion externalities as more products compete for scarce visible store slots, limiting the visibility of, and consumer benefits from, additional varieties.

This paper empirically examines how changes in online consumer discovery costs affect product variety, quality and consumer welfare. I use new data on the Google Play (Android) mobile app store, an online market with over 100 million US consumers and mostly free products/apps. The Google Play store is representative of many online markets. Thousands of new apps appear every week and discoverability is a major concern (iMediaConnection.com). I take advantage of a natural experiment: a re-categorization of apps in the game section of the store.

Re-categorization affects consumer discovery through two channels: (i) the number of game categories increases, reducing the number of apps per category and congestion externalities as fewer apps compete for scarce consumer attention. (ii) category titles become more informative about the apps they contain. Using a difference-in-differences approach I show that discovery costs fall by comparing short-run changes in the downloads of more affected games to less affected games and non-games. I also show that re-categorization increases game entry but reduces the quality of new games.

The relative magnitudes of the various welfare effects resulting from re-categorization are unclear. Consumers should immediately benefit from reductions in discovery costs. If consumers like variety and quality, their welfare should also change in the long-run. Greater variety should increase welfare but reductions in entrant quality and increased congestion externalities mitigate welfare gains from new apps.

I set-up an app demand model and estimate its parameters to evaluate the overall welfare effect of the re-categorization and decompose it. I isolate the immediate welfare gains of better discovery technology while keeping the set of products fixed. I also isolate the long-run welfare channels of increased variety, increased

[1] This is one of the key arguments by the EU Commission in fining Google for restricting visibility of products in its search pages (Europa.eu).

congestion and reduction in entrants' quality relative to a counterfactual market where re-categorization does not happen. Welfare increases by $0.085 per-month per-consumer, which adds up to nearly $100 million per-year in additional welfare across all US Google Play consumers. Around 50% of gains come from immediate improvements in discovery technology for a given set of products. In the longer run each consumer gains $0.087 per-month from greater variety, but higher congestion and reduced entrant quality dissipate around 50% of these welfare gains. Aggregate welfare losses from higher congestion and reduced quality are economically significant, adding up to $44 million and $7 million per-year across all US Google Play Store consumers, respectively.

Table 1: **Google Play Game Categories Before and After March 2014**

Before :
Arcade & Action, Brain & Puzzle, Card & Casino, Casual, Racing, Sports
After :
Action, Adventure, Arcade, Board, Card, Casino, Casual, Education, Family, Music, Puzzle, Racing, Role Playing, Simulation, Sports, Strategy, Trivia, Word

My data consists of weekly and monthly snapshots of all apps available in the US Google Play store from January 2012 to December 2014. Multiple surveys during the sample period show that app consumers primarily discover products by browsing through platform-defined categories (e.g., "Productivity Apps").[2] Many other online markets such as eBay, Amazon and Netflix are organized similarly. I take advantage of a re-design of part of the Google Play store. App stores are split into "game" and "non-game" sections. In March 2014, Google Play re-categorized its 6 game categories into 18 categories (Table 1).[3] Industry observers and anecdotal consumer responses suggest that re-categorization improved consumer discovery.[4]

Before re-categorization the category structure was uninformative and difficult to browse. Different games were grouped together: the Card & Casino category included slot-machine, blackjack, solitaire and Uno-type games. A consumer looking for an Uno-type game would have had to browse through many irrelevant slot-machine games. Other types of games (e.g., Family games) had no dedicated category-space

[2]See Section 2.1 and Online Appendix A.3.

[3]The 24 non-game categories do not change and are listed in Table A1 in Online Appendix A.

[4]See Figure A1 in Online Appendix A.2 and (AndroidCommunity.com).

and there was no simple way for consumers to find them.[5] Re-categorization increased the number of categories, which made it easier for consumers to browse and discover additional games.[6] Re-categorization also made the category structure more informative, giving app-types such as Family games their own areas.

This policy explicitly aimed to reduce consumer discovery costs but its timing was relatively surprising (see Section 2.3). As non-game categories were not changed, I first compare game and non-game outcomes using a difference-in-differences approach. I show that both changes to category informativeness and the number of apps per category affected downloads. Downloads of games that did not have categories before the change (e.g., Music games) increase more than downloads for games that had well-defined categories (e.g., Racing games). I also show a causal negative relationship between the number of apps in a category and app downloads by comparing changes in downloads of apps that were split off from the same category: Card and Casino apps, and Arcade and Action apps. There were more Card than Casino apps, and more Arcade than Action apps, and I show that smaller types disproportionately benefit from the split. This points to the presence of congestion effects, as apps in categories with fewer other apps are more likely to be visible to consumers in either best-seller or other featured category lists.

On the supply side, I show that entry, the number of new unique apps that appear each month, increased by 34% for games relative to non-games after re-categorization. The quality of new games - as measured by consumer ratings and other metrics - fell. I also show supply-side effects are driven by games with greater decreases in discovery frictions.

I set up a structural demand model to measure and decompose the welfare effects of re-categorization. I do not explicitly model the many possible ways in which consumers can search the app store.[7] Instead, I include controls that capture the main sources of discovery friction in this market, and that are consistent with a broad set of implications about consumer search from theory models. App-type fixed effects that vary before and after re-categorization capture changes in informativeness: e.g., the average change in consumer utility from downloading a Music game before as compared to after re-categorization. I verify these are driven by the re-categorization

[5] As discussed in Section 2.1, the search function in the Google Play Store was difficult to use.

[6] Increasing the number of categories should not *always* reduce discovery costs. At some point, having too many categories can result in a more difficult discovery process. There is likely an optimal number of categories to display in this market, but this paper cannot say what this number may be.

[7] In the Online Appendix I also take a more micro-founded approach and set up a simultaneous search and demand model. The model in the main text can be seen as a simplified and linearized version of that model. See additional discussion in Online Appendix D.1.

as opposed to other app-type time-varying demand shocks. I also include apps' past downloads (popularity) and the number of other products in an app's category.[8]

I use the demand model estimates to measure and decompose changes in consumer welfare. I isolate the welfare effect of only changing platform design by calculating expected consumer utility before and after re-categorization with the same set of products. I also calculate expected consumer welfare under simulated market configurations to isolate longer-run changes from product assortment. Reduced form treatment effects provide information about counterfactual market configurations - the number and type of apps that should be available on the store absent re-categorization. For example, entry treatment effects provide information about the counterfactual number of game apps that would have entered absent a re-categorization. I isolate the welfare effect of an increase in the number of products and congestion relative to a market without re-categorization by removing entrants from the factual observed choice-set to match entry treatment effects. I similarly isolate the welfare effects of quality.

I find that consumer welfare increases from the re-categorization, and that over 50% of gains come from immediate changes in discovery costs, rather than from additional entry. Entry does not contribute more because it comes with growing market congestion and reductions in entrant quality relative to a counterfactual market without re-categorization. Increasing congestion and lower entrant quality dissipate over 50% of gains from greater product variety.

This paper contributes to a large literature on the effects of consumer search/discovery costs on competition and market outcomes - starting with Stigler (1961) and Diamond (1971). Most existing theoretical literature focuses on the effects of search costs on price competition; only a small number of recent papers consider non-price effects such as quality and product design.[9] The empirical literature also focuses on the effects of discovery costs and platform design on price competition and match quality.[10] There are few empirical studies on consumer discovery costs and non-price supply-side responses such as entry but they are primarily descriptive.[11] None of these papers separate changes in search costs from other changes in technology that can affect product assortment. My findings suggest that discovery cost reductions were likely responsible for much of online product proliferation.

[8]Apps' past downloads affect positions in best-seller listings that make them more discoverable to consumers. See additional discussion in Sections 2.1 and 5.

[9]e.g., Bar-Isaac et al. (2012), Fershtman et al. (2018), Goldmanis et al. (2010), Moraga-González and Sun (2020).

[10]Fradkin (2017) and Dinerstein et al. (2018).

[11]Bryniolfsson et al. (2011), Zentner et al. (2013).

This paper has two main contributions. No other paper, to my knowledge, uses a plausibly exogenous change in discovery technology to identify the effects of changes in consumer discovery costs on product entry and quality. Product assortment measured by the number of products and their quality is the key competitive outcome of concern for regulators in most online markets.[12] The main policy levers available to regulators of these markets relate to platform design. Quality is important to consumers in online markets and many consumers explicitly look for high quality products (Coey et al. 2020). My results show that platform re-design has a large and persistent effect on the number and type of products available. My results also serve as a valuable test of theoretical predictions (Fishman and Levy 2015, Bar-Isaac et al. 2012, Moraga-González and Sun 2020) and shed light on previous studies that compare offline and online entry. In addition, my findings suggest that platforms can substantially benefit from improvements in discovery technology through increases in downloads/sales.

I also provide the first evidence on the direct and indirect welfare effects of changing consumer product discovery costs in online markets. Falling discovery costs affect welfare directly and immediately. They also affect welfare indirectly by creating changes in product assortment. Increasing variety benefits consumers but generates greater congestion externalities that reduce welfare relative to a counterfactual market where discovery costs do not fall. I show that welfare losses from greater congestion externalities and changing entrant quality are economically meaningful. Consumers still benefit from increasing variety in this market, but it need not be the case in other settings. Such evidence is of direct relevance to policy-makers.

The paper proceeds as follows: Section 2 provides an overview of mobile app market focusing on users (Section 2.1), developers (Section 2.2) and the re-categorization event (Section 2.3). Section 3 describes the data and presents some summary statistics. Section 4 presents the reduced form results. Section 5 discusses the specification and estimation of the structural model and the welfare decomposition. The final section concludes.

[12] In many markets prices are zero or uniform, so changes in competition intensity, platform design, or other market conditions cannot have an effect on prices.

2 App Market Background

2.1 Users

The Google Play Store comes pre-installed on most Android OS phones, making it the most popular app store in the world. Forrester and Nielsen surveys conducted during my sample period show that consumers primarily discover new apps by browsing through the store (see Figure A2 in the Online Appendix).[13] 58% of Android consumers discover new products through "General browsing in the app store", according to Forrester, and 25% discover new products through more targeted browsing - looking at "top rated" or "most popular" app lists in the app store. Only a small share of consumers discover new apps through an internet search engine. Answers are similar in the Nielsen survey, suggesting that store re-designs should have an effect on how consumers discover new products.[14]

Figure 1 shows what consumers see when they browse the Google Play store and the re-categorization that took place in early 2014. The left panel shows the first screen of the Google Play store in 2014. When a consumer opens the Google Play store on their phone, their first choice is between games or non-game apps. After this, consumers choose a category to look at more specific game or non-game product types. The middle panel shows the choice of game categories in late 2013 before the re-categorization took place. The right panel shows the choice of game categories in mid 2014 after the re-categorization took place. Rather than 6 categories (in addition to the "widget" and "live-wallpaper" categories),[15] Google Play re-organized games into 18 different categories in March 2014 (see Table 1). I discuss this event in more detail in Section 2.3.

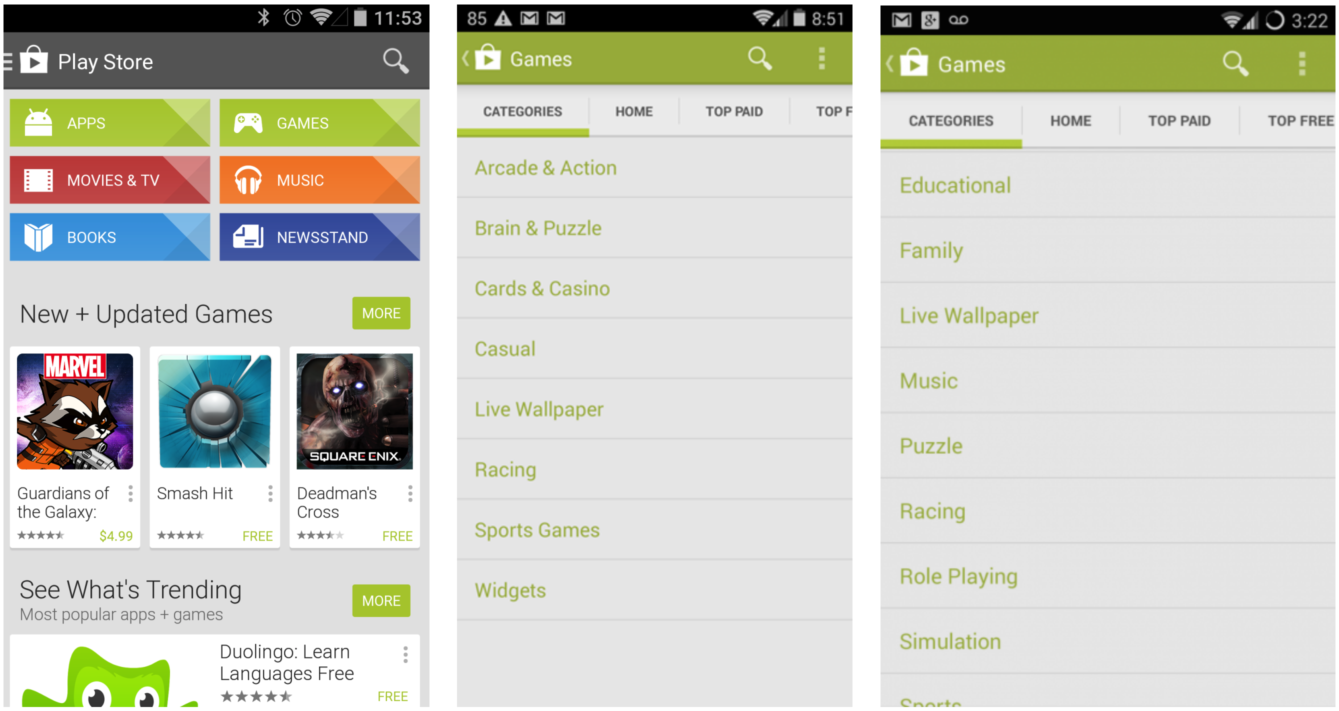
Once a user chooses a category they have several lists of apps to choose from. Bestseller-lists display apps with the largest number of downloads over a certain pe-

[13] Figure A3 shows similar results from additional surveys.

[14] "Physical word of mouth [...]" is the only discovery method that is comparably popular. This discovery channel is not directly affected by the re-design of the store, but there could be some indirect effects. A re-design that changes how consumers browse and what they download could affect word-of-mouth recommendations. The two effects should go in the same direction and estimates of the overall impact incorporate both effects. However, assuming that not all browsing consumers recommend the apps they download, changes in word-of-mouth recommendations would be smaller than changes in direct browsing behaviour, making word-of-mouth effects second order. This is also the case on the supply side: it is easy for app developers to forecast how changes in categorization affect consumer browsing. It is much harder for them to forecast how such changes would further influence consumer word-of-mouth recommendations.

[15] "Widgets" and "live-wallpapers" are not full apps and were generally not particularly popular with consumers. I omit these from my analysis.

Figure 1: Google Play Store Design (c. 2014)



riod of time.[16] Other lists are *not* organized based on past popularity but showcase "featured" apps selected by human editors or algorithmically. There are also lists uniquely dedicated to showcasing new apps. Unlike bestseller-lists which often feature the same apps for long periods of time, apps in other lists within a category rotate frequently. During my sample period consumers *cannot* filter (or selectively browse) categories by app ratings, number of downloads, or other app characteristics. As well, during my sample period the search function in the Google Play Store was unreliable (see additional discussion in Section 2.3). Before downloading an app, consumers observe a number of screenshots from the app, its average rating, how many people have downloaded this app, the size of the app in MB, and a text describing the app.

[16] The exact algorithm determining the position of an app in the top lists is unknown, but it is related to downloads (Adweek.com).

2.2 Developers

The costs of publishing an app on the Google Play market during my sample period are zero - each developer has to pay a one time fee of $25 to register with Google Play and then can publish apps for free (TechRepublic.com). The cost of developing a simple app can be as low as a few thousand dollars. More complex apps that link up to databases (e.g., calendars) can cost up to $50,000 (ApplicoInc.com). At the high end, complex apps (e.g., 3D games, video chatting apps) can cost hundreds of thousands of dollars to develop (SavvyApps.com). Instagram spent over $500,000 to produce a beta version (Instagram Tumblr). Development times are generally short. Estimates suggest that 25% of apps take one month to develop and 75% of apps are developed within six months (Appinventiv.com).

Because of how consumers browse the store (see Section 2.1, an app's category is a key choice by developers. Although it is possible to invest in advertising or otherwise encourage consumers to download an app, success depends on visibility in the store. Apps are visible either by appearing in a category-specific best-seller list, or in one of the other "featured" category-specific lists. Industry analysts suggest that "the one basic rule for choosing categories [is]: find the least competition...any chance to make the process of getting into the top 10 easier should be taken" (PrioriData.com). Category choice is mutually exclusive on Google Play and developers can only choose one category at a time.[17]

Developers earn money in one of three ways. They can charge consumers upfront for downloading their app. These are the "paid apps" and they constitute about 20% of all apps on Google Play. Mean paid app price is $3.3 (see Table 2). Developers can also allow users to download their apps for free and place ads in the app. There is a competitive market for advertisers who bid in second price auctions to display their ads in apps.[18] As in Google's search advertising system (Varian 2007), app ad revenues depend on their popularity. Anecdotal evidence suggests that the average price paid for 1,000 shown ads ("impressions") in the US is approximately $2-3 (Quora.com). Developers can also allow users to download their apps for free, but sell them features or products within the app ("in-app purchases"). Examples of in-app purchases include additional levels for games or magazine subscriptions.

Google makes money from apps on its store. It receives $25 from every new developer. Google also takes 30% of paid app revenues and 30% of in-app-purchases made on free apps (Google). About 50% of ad-supported apps use Google's advertising

[17] After the re-categorization, all game developers with existing apps choose which new category their apps belong to. There is little evidence of category switching otherwise.

[18] This side of the market is beyond the scope of this paper and is not explicitly modelled.

platform, Admob, meaning that Google takes a portion of advertising revenues (AppBrain.com). Even if Google does not earn money from other advertising platforms, it collects app data that helps optimize search and advertising.

2.3 March 2014 Game Re-Categorization Event

The Google Play Store initially had six game categories compared to eighteen game categories in the Apple iOS store. Anecdotal reports suggest that Google was less concerned with product discovery as its store was relatively smaller. Google also expected consumers to primarily use the search bar within the store or for Android apps to be quickly integrated into Google Search results. This was not the case by 2014. The search function in the Google Play Store was unreliable (AndroidAuthority.com). Google introduced app indexing in their Google Search only in early 2015 (after my sample period) but this functionality was not widely adopted by app developers. It required substantial code adjustments and also frequently failed to provide working links to the Google Play Store (SmashingMagazine.com).

On December 9, 2013 Google announced that it was expanding the number of game categories in the first quarter of 2014 (droid-life.com). The new classification matched Apple's existing categorization. Developers with existing apps could choose their own new categories ahead of time. The date for the re-categorization was announced to be February 2014, but Google eventually delayed launching the new categories until March 17, 2014. Industry observers and developers were reportedly surprised both by the initial announcement and by launch delays (AndroidPolice.com).

Re-categorization produced two substantial changes to the consumer discovery process. First, it tripled the number of new game categories and immediately reduced the number of apps in each game category. As discussed in Sections 2.1 and 2.2, categories are important to the consumer search process and the number of apps per category is important to an app's visibility. Tripling the number of categories effectively reduces competition for scarce visible spaces in the store. An app is more likely to feature in the rotating "featured" app lists and be visible to consumers. Second, the titles of categories became much more informative about the types of apps present. Before the new categories, consumers looking for music, family or strategy games did not know precisely where to look. After re-categorization, this changed. Consumers had clear information about where different types of apps were located in the store. On the supply side, developers knew that if they produce such a game, there is a clear place for it to be discovered.[19]

[19]This is especially the case since the new categorization structure already existed on the Apple

These changes are consistent with Google's primary motivation for the re-categorization: improving the consumer search experience. An industry observer describes that "searching for general apps on the Play Store is an exercise in frustration" under the old six game category system (AndroidPolice.com). In the PR announcement for the 2014 change, Google states that the new game categories "mak[e] it easier for players to find games they'll love" (AndroidPolice.com). A Google blog post discussing a subsequent re-categorization in 2016 clearly states that new app categories "improve the overall search experience" and "mak[e] them more comprehensive and relevant to what users are looking for today" (GoogleBlog.com). For the 2014 game re-categorization studied here, the wholesale adoption of Apple's game categories suggests that Google did not choose categories based on pre-existing trends.[20]

The 2014 re-categorization appears to have been successful for Google as evidenced by subsequent category expansions. In 2015 Google introduced additional "Family Game" sub-categories to make it easier for parents to find games for kids of varying ages (AndroidCommunity.com). Google again introduced new categories in 2016 for non-game apps. (GoogleBlog.com). The Apple App store also changed their categories, experimenting with removing some non-game app categories in 2018 (MacObserver.com). These events fall outside the scope of my data. Re-designs also happen in other online markets with similar goals. Amazon changes its product categories frequently and eBay also experimented with re-design (Dinerstein et al. 2018).

3 Data

3.1 Data Description

My data comes from AppMonsta.com and consists of daily snapshots of all apps on the US Google Play store, aggregated at the weekly or monthly level, starting from January, 2012 and up to December, 2014.[21] This is the first paper to use this dataset.[22] The data contains all information that consumers observe on the Play

store for years at that point and developers frequently produce apps for both platforms.

[20]Formal tests for downloads in Figure C1 and Online Appendix C.2.1, and for entry and quality in Figure C3 show sharp changes in outcomes exactly around the period of re-catgeorization but no differences before the policy's announcement. Private discussions with Google employees familiar with changes to the Play Store also confirm that app entry is not a consideration when making such changes. They suggest opposite concerns about too many products already appearing in the store.

[21]Weekly aggregation is only used to predict app downloads (see Section B.2).

[22]Liu et al. (2014) use an app dataset from the same provider for 2011-2012.

store - app price (in USD), a histogram of the ratings the app received (ranging from 1 to 5), app size (in MB), the number of preview screenshots the app shows, the number of video previews the app shows, and a download range for the app (number of lifetime downloads). I also observe the app's category, the name of the app's developer, and a text description of the app.

I also scraped historical app rankings from Flurry.com. With this data I observe the "top lists" for every category in each week, which approximate the top 500 weekly best-selling free and paid apps in each category.[23]

3.2 Data Management

3.2.1 App-Type Classification

I classify game apps into types based on their categories in the last period of the sample.[24] An app belonging to the Music category in the last period after re-categorization is defined to be a "Music" type game.[25] Table B4 in the Online Appendix shows some summary statistics at the app-type level. There are 42 app-types, 18 of which are game types. Game app-types have fewer apps than non-game types. The average game type is less than a quarter of the size of the average non-game type.

Game app-types, denoted by c for the rest of the paper, are distinct from categories, denoted by c^*. Before re-categorization, multiple app-types can be present in one category. Before re-categorization, app-types can also be present in categories whose names do not represent them. For example, Music or Family type games can be in the Arcade & Action category, or in the Brain & Puzzle category. For non-games, the distinction between categories and app-types does not practically matter

[23]See additional discussion about "top lists" and bestselling lists in Online Appendix B.2.

[24]Apps can also be classified into a larger number of types based on alternative criteria, as in Kesler et al. (2020). However, it is likely that app developers think about their apps in terms of the 18 post re-categorization types because the Apple App Store already had that category structure for years and developers often multi-home.

[25]There are two possible drawbacks to this approach: (i) Approximately 1% of games exit the market before re-categorization and cannot be classified in this way. I use app descriptions to classify *only* these apps into post- re-categorization types. See Online Appendix B.1 for a description of this approach. An alternative approach that drops these apps does not change the results. (ii) There is possibly some selection of apps into categories based on competition and other features of the market. i.e., it is possible that a "Music" type app enters into another category. However, both of these concerns are likely minor in practice. Previous versions of the paper used a machine learning algorithm and text-based analysis to classify apps into types where concerns (i) and (ii) apply less. Results are quantitatively similar across the two classification approaches.

as fewer than 0.2% of apps change categories in my sample.[26]

3.2.2 Predicted Downloads

In the raw data I do not observe apps' weekly or monthly downloads but only lifetime download bandwidths reflecting how many downloads an app has had throughout its entire time in the store.[27] These can range to millions of downloads.[28] I estimate monthly downloads using information about app rankings in each category and week and the download bandwidth of new apps (apps that arrived in the market in that week). I recover a relationship between rankings and downloads for new apps.[29] Then, I predict the downloads of all other apps in the market. Section B.2 in the Online Appendix provides more details about this procedure.[30]

This approach relies on a functional form assumption for the relationship between downloads and rankings and could produce inaccurate estimates of monthly downloads (Liebowitz and Zentner 2020). As a robustness check, I use an alternative proxy for downloads: the difference in the number of ratings an app receives between two consecutive months. Results are qualitatively similar across the two approaches. See additional discussion in Online Appendix C.2.3.

3.3 Descriptive Statistics

Table 2 shows some summary statistics at the app level. There are approximately 33.7 million app-month observations in total, consisting of 2.6 million unique apps. Of these, approximately 17% belong to game categories. 20% of apps have non-zero prices. The average price for a paid app is approximately \$3.3. I provide additional

[26] This is likely because developers are very careful in deciding on the initial category positioning. Re-positioning is risky as it moves an app to compete with a new set of other apps for scarce consumer attention and could push it down a bestseller list or make the app less likely to be "featured" on the store otherwise.

[27] These downloads do not include updates or double count the downloads of different versions of the app.

[28] Table B1 in Online Appendix B.2 shows all download bandwidths.

[29] Intuitively, I observe a new app ranked 1st in a given category with a lower bound of 50,000 download, a new app ranked 10th with a lower bound of 10,000 downloads, and a new app ranked 100th with a lower bound of 500 downloads. Under certain distributional assumptions, I can recover the relationship between the lower bound of downloads and ranking. See additional discussion in Online Appendix B.2.

[30] I predict zero downloads for about 20% of apps in a given period. To mitigate the "zeroes" problem in demand estimation, under some assumptions about the underlying distribution of consumer downloads I apply the method of Gandhi et al. (2014).

Table 2: **App Summary Statistics**

Variable	Mean	Std. Dev.	N
App Level			
Game App Dummy	0.168	0.374	2.6 million
Paid App Dummy	0.2	0.4	2.6 million
App-Month Level			
Lifetime Downloads (Min.)	38,261	1.9 million	33.7 million
App Size (in MB)	21.99	29.75	33.7 million
Monthly Predicted Downloads	559	25,169	33.7 million
Number of Screenshots	4.71	3.54	33.7 million
Number of Videos	0.09	0.28	33.7 million
Mean Rating	4.0	0.66	33.7 million
Price (for Paid Apps)	3.27	8.93	6.8 million
App-Month Level for Section 5 Sample: Game-Apps			
Lifetime Downloads (Min.)	64,269	904,779	4,152,147
App Size (in MB)	11.645	42.982	4,152,147
Monthly Predicted Downloads	236	3,509	4,152,147
Number of Screenshots	5.827	4.564	4,152,147
Number of Videos	.176	.38	4,152,147
Mean Rating	3.355	1.618	4,152,147
Price (for Paid Apps)	1.92	4.754	796,522

summary statistics for the sub-sample of game apps I use in Section 5 to estimate the demand model. Average price for paid games is approximately $1.9.

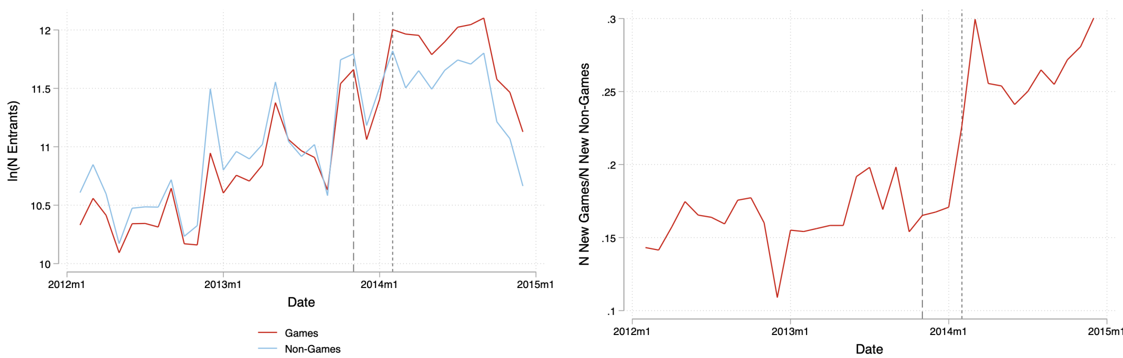
4 Reduced Form Evidence

In this section I test whether the re-categorization of games affected consumer discovery costs, as proxied by downloads, and developer entry and quality decisions. I also separately show the two key mechanism through which re-categorization affects consumer and firm decisions: category titles became more informative, and the number of apps per category fell. I use a difference-in-differences identification strategy, assuming that re-categorization treats game apps and non-game apps serve as a control group. This is a reasonable approach given the surprising timing of policy and the separation between games and non-games in the store (see Section 2.3).[31]

Figure 2 provides an intuitive justification for using difference-in-differences in this setting, showing aggregate monthly entry for games and non-game apps. The left hand panel shows the monthly number of new games and non-games after removing average differences between groups. The right hand panel shows a ratio of the number of new games relative to the number of new non-games.

Figure 2: Google Play Entry Patterns

(a) Monthly Entry (b) Monthly Entry Ratio



Notes: Each line in panel (a) shows the log of the total number of new game or non-game apps appearing in the Google Play Store in a month. Both time series have been de-meaned to remove average differences between games and non-games. Seasonality in the form of month fixed effects has also been removed from each time series in panel (a). Panel (b) shows a ratio of the monthly number of new games over the monthly number of new non-games appearing in the Google Play store in a month ($\frac{\text{N New Games}}{\text{N New Non-Games}}$). In both panels, the first dashed vertical line represents the re-categorization announcement and the second dashed vertical line represents the re-categorization period.

The left hand panel shows substantial changes in absolute entry for games and non-games over the sample period.[32] But the right hand panel shows that *relative* game and non-game entry in 2012 and 2013 is nearly constant. This changes two months after the announcement of the game re-categorization in late December 2013. Starting in February 2014 there is an increase in entry by game apps but not by non-game apps. The increase in entry peaks after re-categorization takes place in March

[31] I formally test for parallel trends for downloads in Online Appendix C.2.1 and Figure C1. Formal timing tests for entry and quality are in Figure C3 in the Online Appendix.

[32] There are many possible reasons for this variation, such as the release of new phones or OS updates. An increase in entry in mid-2013 coincides with the release of the Samsung Galaxy S4 - a flagship Android phone which became the best-selling smartphone in the world (GSMArena.com).

2014 and remains nearly constant after that. This suggests that re-categorization increases entry by games.

More formally, an outcome (i.e., number of new apps, average quality) for app type c at time t is:

$$y_{ct} = \tau(Game_c \times Post_t) + \delta_c + \delta_t + \epsilon_{ct} \tag{1}$$

where δ_c and δ_t are app-type and time fixed effects, $Post_t$ is a dummy equal to one after re-categorization from March 2014, and $Game_c$ is a dummy variable equal to one for the game types and zero for non-game types. The coefficient of interest in Equation 1 is τ, which captures the treatment effect of re-categorization on outcome Y.

I estimate Equation 1 at two additional aggregation levels. I add up or average outcomes across all games and non-games such that I have one game and one non-game observation per month. This aggregation level is most comparable to Figure 2. I also estimate Equation 1 at a more dis-aggregate individual app-level with app fixed effects.

4.1 Downloads

Re-categorization affected two of the main source of discovery frictions for consumers browsing through categories on the Google Play store. First, it made the category structure more informative, making it more obvious for consumers where to find music, family or educational games. Second, by introducing more categories, it reduced the number of apps per category and changed the probability an app becomes visible to consumers. In traditional retail spaces, finite physical shelf-space for products limits the number of varieties available to consumers. Shelf-space is in principle unlimited online, but in practice products need to be prominently displayed on best-seller or other lists of featured products to be visible (see Section 2.1).[33] The more products exist in a category space, the more products are competing for scarce visible slots, creating congestion externalities and crowding out other products online just as in offline retail.[34]

[33] An app's past success affects positioning in category best-seller rankings and it is an additional source of discovery friction. I do not focus on this friction in the reduced form section as this is not directly affected by Google Play's re-categorization. Results, available on request, show that an app's past success has a direct effect on its downloads in the present, conditional on quality. The model in Section 5 captures this source of search friction directly.

[34] I discuss the possibility that changing the number of apps in a category affects competition intensity instead of congestion below. Results in this Section and in Online Appendix C.10 suggest this is unlikely.

I test for both effects in the data. To capture the direct effects of re-categorization on consumer discovery costs and demand, I focus on app downloads at the aggregate and at the app level. I do not observe data on app usage.[35] Theory predicts that reductions in product discovery/search costs should improve consumer/product match values and increase downloads (i.e., Weitzman 1979). I use the predicted monthly app downloads described in Section 3.2.2.[36]

Before focusing on the two discovery friction channels, I test whether game downloads increase on average relative to non-game downloads after re-categorization. I do this for all levels of data aggregation: at the game/non-game level, at the app-type level and at the individual app level. I first estimate Equation 1 using the full January 2012-December 2014 time period. However, product assortment in the app market is constantly changing through entry and re-categorization affected entry patterns. To minimize confounding changes in product assortment and category informativeness, I re-estimate Equation 1 for a narrow four month time period: the two months immediately before re-categorization (January and February 2014), and the two months immediately after re-categorization (March and April 2014).[37] I consider the narrow time sample as my main specification for download-related outcomes. I also test for non-existent re-categorization events before and after the actual re-categorization in Online Appendix C.2.1.[38]

Estimates are in Table 3. Columns (1)-(3) show results for the full 2012-2014 sample, and columns (4) and (5) show results for the narrow January-April 2014 sample. They show that game downloads increase relative to non-game downloads after re-categorization. At the most aggregate level in Column (1), downloads in-

[35]Downloads may be more correlated with usage for paid apps than for free apps. I estimate effects separately for paid apps online in Online Appendix C.2.2. Results are qualitatively and quantitatively similar to those in the main text.

[36]This measure of app downloads requires making functional form assumptions on the relationship between downloads and app rankings. I also use an alternative measure of downloads that does not require these assumptions. Results are in Online Appendix C.2.3 and are qualitatively similar to those in the main text.

[37]I focus on four months rather than just February and March since the re-categorization happened roughly in the middle of March so effects identified using February and March are likely incomplete. I also show that the main results hold for a shorter two month February/March time period in Table C1 in the Online Appendix. As anticipated, effects are qualitatively similar but quantitatively smaller than for the four month sample.

[38]Using November 2013-February 2014 data , I test for a non-existent re-categorization event between December and January. Using March 2014-June 2014 data I test for a non-existent re-categorization event between April and May. Estimates are in Table C2 and show no statistically significant changes for both "placebo" events. For the "full" time period regressions (January 2012 - December 2014), I allow treatment effects to vary over time and similarly show no changes before or after re-categorization in Figure C1.

Table 3: Downloads Difference in Differences Estimates: Average Effects

Outcome Variable:	ln(Tot. Downloads) (1)	ln(Tot. Type Downloads) (2)	ln(App Downloads) (3)	ln(Tot. Type Downloads) (4)	ln(App Downloads) (5)
Games × Post	0.748***	1.275***	0.162*	1.399**	0.438*
	(0.112)	(0.203)	(0.090)	(0.288)	(0.167)
Games	-7.423***				
	(2.860)				
Unit of Observation:	Agg. Game/Non-Game	App-Type	App	App-Type	App
Time Period:	Jan 12/Dec 14	Jan 12/Dec 14	Jan 12/Dec 14	Jan 14/Apr 14	Jan 14/Apr 14
Sample:	All	All	All	All	All
Year/Month FE	•	•	•	•	•
App-Type FE		•		•	
App FE			•		•
App Controls			•		•
Observations	70	1,470	32,964,682	168	5,284,311
R-squared	0.970	0.929	0.953	0.780	0.972

Notes: The sample period in the first three columns is January 2012-December 2014 and in the last two columns is January 2014-April 2014. All app types and apps are considered in each sample. Data in Column (1) consists of monthly observations at the Game/Non-Game level. Data in Columns (2) and (4) consists of monthly observations at the app-type level. Data in Columns (3) and (5) consists of monthly observations at the app level. Outcomes are natural logarithms of downloads at each aggregation level. Controls include year and month fixed effects, game/non-game fixed effects, app-type fixed effects, or app fixed effects, depending on the column. Additional controls in Columns (1)-(3) include game/non-game or app-type specific time trends. Additional app-level controls for Columns (3) and (5) include average app ratings, a dummy for whether the app is free or paid, the price of the app if it is paid, and app age-specific fixed effects. The variable "Games × Post" is a dummy variable equal to 1 for games, or game app-types/apps starting from March 2014. Standard errors are robust to heteroskedasticity in Column (1) and are clustered at the app-type level in the remaining columns. *** p<0.01, ** p<0.05, * p<0.1

crease by 75%.[39] At the app-type level, this corresponds to an increase of nearly 130%. Individual monthly app downloads increase by 16%.[40] Estimates are also similar in regressions using only short run variation in Columns (4) and (5). These results are suggestive of the re-categorization changing consumer download patterns but they do not necessarily show evidence of changes in discovery costs. Next, I show evidence for the two channels of reduced consumer discovery frictions.

First, I focus on the informativeness of the category structure. Eight of the eighteen app-types had visibility as game categories before the change: Action and Arcade were grouped together as Arcade & Action, Card and Casino were grouped as Card & Casino. Puzzle was named Brain & Puzzle, and Racing, Sports and Casual

[39] Panel (d) of Table C4 shows estimates with absolute (non-log) downloads. They correspond to an increase of approximately 2 million downloads per month.

[40] Aggregate estimates are larger than per-app estimates for two reasons: (i) aggregate estimates capture both changes in product assortment and changes in per-app downloads, whereas the app estimates only capture the latter. (ii) As I explain below, some app-types are more affected by re-categorization than others. Less affected app-types also generally have more apps. In the aggregate regressions each app-type has equal weight, whereas in the individual app regressions app types with more apps have greater weight in the average estimated effects.

games remained unchanged. The remaining app-types did not have pre-existing categories: Adventure, Board, Education, Family, Music, Role Playing, Simulation, Strategy, Trivia and Word games.[41] Consumers should have become much better informed about the location of these app-types after re-categorization and should be able to reach them much faster. If category informativeness plays a role in discovery frictions, downloads for app-types without pre-existing categories should be more affected by re-categorization than app-types with pre-existing categories. I estimate the following regression at the app-type and app level:

$$y_{(j)ct} = \tau^1 \text{Post}_t \times \text{Game}_c + \tau^2 \text{Post}_t \times \text{Game}_c \times \text{No Pre-Existing}_c + \delta_{(j)c} + \delta_t + e_{(j)ct} \quad (2)$$

This regression is estimated using the four months around the re-categorization event (January-April 2014).[42] δ_t and $\delta_{(j)c}$ are month and app-type or app fixed effects, Post_t is a dummy equal to one in March and April 2014, Game_c is a dummy equal to one for all app-types or apps that are games, and No Pre-Existing_c is a dummy equal to one for the ten app-types that did not have categories before March 2014 and zero otherwise. Non-game app-types/apps are the baseline group.[43]

Results from these regressions are in panel (a) of Table 4. They show that app-type-level and app-level downloads increase more for games that did not have pre-existing categories. There is a 44 percent increase in downloads after re-categorization for app-types with pre-existing categories, but the change for an app-types without pre-existing categories is four times as large. The heterogeneity is similar at the app level after controlling for app fixed effects. In Columns (2) and (4), I restrict the sample by excluding some game and non-game types that are very different than game types without pre-existing categories.[44] I still find similar heterogeneity in effects. In Table C2 I also show statistically null effects in response to non-existent events taking place before and after actual re-categorization. These results suggest

[41]Even though they did not exist in the Google Play store, these were categories in the Apple iOS app store for several years prior.

[42]I also estimate it using only February and March 2014 in Table C1. Results are qualitatively similar but quantitatively smaller.

[43]It is possible that there are some informativeness effects even for app-types with pre-existing categories. For example, consumers searching for Casino apps know after re-categorization that there are only Casino-type apps in the Casino category. Apps belonging to these types also experience changes in the number of other apps in their categories and congestion, which also affects downloads as I show below.

[44]Apps without pre-existing categories have fewer apps on average than game types with pre-existing categories or non-game types. I exclude the largest non-game and game types (by the mean number of apps in 2013) to address this concern.

that re-categorization made the category structure more informative and reduced consumer discovery frictions, increasing downloads.

Second, I test for the effects of re-categorization on congestion coming from the number of apps in a category. I have two distinct empirical strategies to identify this channel. I initially focus on a subset of game app-types with pre-existing categories. Four post-policy game categories are derived from two pre-policy categories: Arcade & Action split into Arcade and Action games, and Card & Casino split into Card and Casino games.[45] Although they originate from the same pre-policy category, the number of apps between the split app-types was unequal before re-categorization. There were three times as many Arcade games compared to Action games, and three times as many Card games as compared to Casino games. Before re-categorization, a Card game and a Casino game had the same number of other apps in their category affecting their chances of being visible to consumers. If congestion exists and the number of apps in a category matters for discovery, re-categorization should benefit Action games more than Arcade games, and Casino games more than Card games. I test whether immediate changes in downloads after re-categorization were larger for the app-types with fewer apps by estimating the following regression:

$$y_{(j)ct} = \tau^1 \text{Post}_t \times \text{Game}_c + \tau^2 \text{Post}_t \times \text{Game}_c \times \text{Small Type}_c + \delta_t + \delta_{(j)c} + e_{(j)ct} \quad (3)$$

where Small Type_c is a dummy equal to one for Action and Casino apps/app-types and zero otherwise. I restrict the sample to include only the four game app-types described above, but I include all non-game app-types as the baseline group. I restrict the sample period to the four months from January 2014 to April 2014 to limit confounding changes in product assortment.[46]

Estimates from this regression at the app-type and app levels are in Columns (5) and (6) in panel (b) of Table 4. They show that downloads increase more after re-categorization for smaller sub-categories/types as compared to larger sub-categories/types. This heterogeneity is substantial. The baseline average increase in downloads for large types is 30%, but downloads increase for small sub-categories by 70-120%. At the app-level, these effects are conditional on app fixed effects and additional app controls. This evidence is consistent with congestion externalities, as the actual number of alternative Card and Casino games is not changing much

[45] Of the remaining game types that existed as categories before re-categorization, Casual, Racing and Sports games did not formally change. Brain & Puzzle transformed into Puzzle games.

[46] Results using only February and March 2014 are in Table C1. They are qualitatively identical but quantitatively smaller than in Table 4. Null estimates for non-existent events before and after actual re-categorization are in Table C2.

Table 4: Downloads Difference in Difference Estimates: Discovery Cost Channels

Panel (a): Category Informativeness

Outcome Variable:	ln(Tot. Type Downloads) (1)	ln(App Downloads) (2)	ln(Tot. Type Downloads) (3)	ln(App Downloads) (4)
Games × Post	0.442*	0.220	0.545**	0.476
	(0.164)	(0.120)	(0.134)	(0.215)
Games × Post × No Pre-Existing	1.723***	1.987***	1.635***	1.665***
	(0.241)	(0.191)	(0.267)	(0.195)
Unit of Observation:	App-Type	App-Type	App	App
Sample Period:	Jan 14/Apr 14	Jan 14/Apr 14	Jan 14/Apr 14	Jan 14/Apr 14
Sample:	All	All	Small Types	Small Types
Year/Month FE	•	•	•	•
App-Type FE	•	•		
App FE			•	•
App Controls			•	•
Observations	168	5,284,311	72	306,956
R-squared	0.916	0.980	0.918	0.935

Panel (b): Number of Apps in a Category / Congestion

Outcome Variable:	ln(Tot. Type Downloads) (5)	ln(App Downloads) (6)	Post/Pre Δ ln(App Downloads) (7)	Post/Pre Δ ln(App Downloads) (8)
Games × Post	0.339*	0.229		
	(0.129)	(0.111)		
Games × Post × Small Type	0.745***	1.028***		
	(0.085)	(0.153)		
Post/Pre Δ ln(N Apps in Category)			-0.651***	-0.650***
			(0.003)	(0.026)
Post/Pre Δ ln(N Apps in Category) × No Pre-Existing				-0.003
				(0.033)
Unit of Observation:	App-Type	App	App	App
Sample Period:	Jan 14/Apr 14	Jan 14/Apr 14	Jan 14/Apr 14	Jan 14/Apr 14
Sample:	All Non-Games + Arcade, Action, Card and Casino	All Non-Games + Arcade, Action, Card and Casino	All Games	All Games
Year/Month FE	•	•		
App-Type FE	•			
App FE		•		
App Controls		•	•	•
Observations	112	4,770,936	142,254	142,254
R-squared	0.953	0.984	0.879	0.879

Notes: The sample period in all columns is January 2014-April 2014. Data in Columns (1),(3) and (5) consists of monthly observations at the app-type level. Data in Columns (2), (4), (6), (7) and (8) consists of monthly observations at the app level. Columns (1) and (2) include all apps. Columns (3) and (4) include all app-types without pre-existing categories and other non-game and game app types with fewer than 20,000 apps in 2012. Columns (5) and (6) include all non-game apps and Arcade, Action, Card and Casino games. Columns (7) and (8) include all game apps. Outcomes for Columns (1)-(6) are the natural logarithms of downloads at each aggregation level. The outcome for Column (7) is the difference between the natural log of average app downloads in March and April 2014 and the natural log of average app downloads in January and Feburary 2014. Controls include year and month fixed effects and app-type or app fixed effects depending on the column. Columns (7) and (8) do not include app fixed effects since they are in first differences. Additional app-level controls include average app ratings, a dummy for whether an app is free or paid, the price of an app if it is paid and app-age specific fixed effects. The variable "Games × Post" is a dummy variable equal to 1 for games (or game app-types for even columns) during and after March 2014. The variable "Games × Post × No Pre-Existing" is a dummy variable equal to 1 during and after March 2014 only for games/app-types that did not have pre-existing categories before March 2014. The variable "Games × Post × Small Type" is a dummy variable equal to 1 during and after March 2014 only for Action and Casino games. Δ ln(N Apps in Category) is the difference in the natural log of the number of apps in the category of app j after re-categorization and the natural log of the number of apps in the category of app j before re-categorization. Standard errors are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

between just before and just after re-categorization. What changes is the number of irrelevant alternatives that compete for scarce visible spaces, which decreases more for Casino than for Card games and more for Action than Arcade games. I show that this change drives changes in downloads.

My second identification strategy takes advantage of the fact that re-categorization reduced the number of apps per category for *all* game categories and not just for categories that were split. Even a category like Racing games whose title was not changed saw reductions in the number of apps, likely because some apps were better described as Music or Family games and fit better into one of the new category spaces. This means that for any game app, the number of other apps in their category declined immediately after re-categorization.[47] For a given app, the differences in the number of other apps in its category immediately after re-categorization as compared to before re-categorization are not driven by entry. As in my first identification strategy, competition is also held relatively constant. I use this to provide evidence relating changes in congestion through the number of apps in a category to downloads.

For each *game* j, I regress the first difference in downloads on the first difference in the number of apps in their category around the period of re-categorization.[48] The first difference absorbs most time-invariant app characteristics, and I also control for other app characteristics like average rating and price. The estimating equation is:

$$\begin{aligned} ln(Downloads_{j,Post}) - ln(Downloads_{j,Pre}) = \alpha(ln(NApps_{jc,Post}) - ln(NApps)_{jc^*,Pre}) \\ + \beta X_j + \epsilon_j \qquad (4) \end{aligned}$$

where $Downloads_{j,Post}$ are game j's average monthly downloads in the two months after re-categorization, and $Downloads_{j,Pre}$ are game j's average monthly downloads in the two months before re-categorization.[49] $NApps_{jc,Post}$ and $NApps_{jc^*,Pre}$ are similarly the number of apps in the category of app j before and after re-categorization.[50]

[47] I show direct evidence of this in Figure C2 in the Online Appendix. For each game app, I compare the number of other games in their category in January and February 2014 to the number of other games in their category in March and April 2014.

[48] I also estimate the same regression with the difference in paid app price as the dependent variable. This regression helps verify that the change in the number of apps in a category operates primarily as a change in consumer discovery costs, rather than a change in direct competition intensity. See Online Appendix C.10 for additional discussion.

[49] In Column (7) of Table C1 I show that the results hold using only February and March data.

[50] Before re-categorization app j's category is represented by c^*. After re-categorization app j's category coincides with its type, c.

I also test whether this relationship varies with changes in the informativeness of category structure by interacting the difference $(ln(NApps_{jc,Post}) - ln(NApps)_{jc^*,Pre})$ with the "No Pre-Existing$_c$" dummy.

Estimates of this regression are in Columns (7) and (8) of Table 4. They show a statistically and economically significant negative relationship between the number of apps in a category and app downloads. A one percent increase in the number of apps in a category reduces app downloads by 0.65 percent.[51] Estimates in Column (8) show that this relationship is independent of changes in the informativeness of the category structure. The average growth rate of the number of apps-per-category in game categories after re-categorization was 9%, suggesting that congestion can have substantial effects on consumer demand. In Online Appendix C.4 I find a similar elasticity of downloads with respect to longer run changes in the number of apps in a category.

In Section 4.2 I show evidence of the effects of re-categorization on product variety and product quality. The presence of congestion raises questions about the welfare gains consumers experience from additional product variety. Although the creation of new category spaces in the app store increases opportunities for discovery conditional on the number of apps, the entry response to re-categorization can reduce the probability that a particular app makes it to a bestseller list or a featured product list. Despite the *existence* of additional products that could satisfy consumers' love of variety, in practice consumers could have access to only a limited subset of these products.

This evidence matches industry analyst statements about the importance of creating products in niches with relatively fewer competitors and maximizing chances of being featured on the store (PrioriData.com). It also reflects consumer complaints about the existence of too many apps in the market and the difficulty of findings products they are looking for (Forbes). I evaluate the importance of congestion effects on consumer welfare in Section 5.

Overall, results in this section show that re-categorization immediately decreased consumer discovery costs in the Google Play store. It did so through two channels: increasing the informativeness of category structure and reducing the number of apps per category and congestion.[52]

[51]This may be a biased estimate of the true effect as apps can strategically sort into categories. However, if that is the case, the main concern comes from apps with better demand shocks or unobservables sorting into "better" categories with fewer other apps. This would suggest a *negative* correlation between the error term and the number of apps in a category, and a *positive* correlation between the error term and downloads. Together, this means I under-estimate (in absolute terms) the effects of the number of apps on app downloads.

[52]I show some additional decomposition of which apps benefit most from the re-categorization

4.2 Entry and Quality

Theory suggests that changes in consumer discovery costs affect supply side firm decisions. When costs fall and firms know they are more likely to be discovered by consumers, there should be an increase in product entry (Anderson and Renault 1999, Chen and Zhang 2017). A reduction in discovery costs can also change product quality, although the direction of this change is ambiguous (Fishman and Levy 2015, Moraga-González and Sun 2020).

I empirically test these predictions using the re-categorization event that reduced discovery costs for game apps relative to non-game apps. For app entry, I observe the number of new unique apps that appear each month.[53] I use consumer ratings to measure quality. In the main text, I show use a quality measure commonly used in the literature: an app's average rating.[54]

I estimate the regression in Equation 1 using the January 2012-December 2014 time period. Since I am looking for aggregate effects, regressions are at the aggregate game/non-game or app-type level. Table 5 shows results for the two main supply-side outcomes. Columns (1) and (2) use monthly entry by new apps as an outcome variable.[55] Columns (3) and (4) use average ratings for the new entrants as an outcome.

Entry treatment coefficients are 0.34 at the game/non-game level and 0.56 at the app-type level. The estimates are statistically significant at the 99% confidence level. They show that following the re-categorization developers entered 34% more game apps than non-game apps. At the app-type level, entry increased by over 50% for game app-types relative to non-game app-types.[56] These results are consistent with

in Online Appendix C.5. Consistent with the structure of the store, I find that older and more previously successful apps benefit more, as they are more likely to appear in the newly created bestseller lists or other featured lists.

[53]I only consider new apps that appear on the store, rather than apps that switch categories or produce new versions. In general, very few apps switch categories. Less than 0.2% of apps switch categories. A fraction of that percentage switch between being classified as games and non-games.

[54]I observe other proxies in the data, including an app's percentage of one star ratings (as in Chevalier and Mayzlin 2006), app size in MB and the information that apps reveal about themselves (e.g., number of photos or video previews). Quality proxies are strongly correlated and produce similar qualitative results. See additional discussion in Online Appendix C.7. In that Appendix I also show that download-weighted average ratings produce similar qualitative results and even larger quantitative results. Ratings-based quality measures are calculated using apps with at least 5 ratings.

[55]This measure only counts apps that have previously not existed in the store and became active. Existing apps that changed categories or that had updates are not counted in this measure.

[56]See Table C9 in the Online Appendix for absolute entry results which estimate that an additional 1,500 game apps enter in the average game app-type after re-categorization.

Table 5: Entry and Quality Difference in Differences Estimates: Average Effects

Outcome Variable:	ln(N Entrants) (1)	ln(N Type Entrants) (2)	Mean Entrant Rating (3)	Mean Type Entrant Rating (4)
Games × Post	0.342***	0.556***	-0.005	-0.061***
	(0.068)	(0.095)	(0.004)	(0.023)
Games	-6.914***		0.253*	
	(1.893)		(0.133)	
Unit of Observation	Game/Non-Game	App-Type	Game/Non-Game	App-Type
Time Period	Jan 12 / Dec 14	Jan 12 / Dec 14	Jan 12 / Dec 14	Jan 12 / Dec 14
Sample	All	All	All	All
Year/Month FE	•	•	•	•
App-Type FE		•		•
Observations	70	1,470	70	1,470
R-squared	0.997	0.975	0.979	0.674

Notes: The sample period in all columns covers January 2012 to December 2014. Sample in odd columns includes monthly observations at the Game/Non-Game level. Sample in even columns includes monthly observations at the app-type level. Column (1) outcome is the natural logarithm of number of new games/non-games on Android. Column (2) outcome is the natural logarithm of the number of new apps in a given game/non-game *app-type* on Android. Column (3) and (4) outcomes are the average ratings of entrants at the game/non-game or app-type level. Controls include year and month fixed effects, game/non-game fixed effects for odd columns, and app-type fixed effects for even columns. Additional controls include app-type specific time trends. The variable "Games × Post" is a dummy variable equal to 1 for games (or game app-types for even columns) during and after March 2014. Standard errors are robust to heteroskedasticity in odd columns and clustered at the app-type level in even columns. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

raw data presented in Figure 2.

Entry changes reflect theoretical predictions from search models. As product discovery costs fall, consumers are more likely to search more and discover new products. The discovery probability for a given product rises and firms respond by entering more products. Coefficient magnitudes are reasonable in the context of a market that is cheap/free to enter and where new products are relatively cheap to produce (see Section 2.2). Results are consistent with previous descriptive comparisons between high and low discovery cost markets. Brynjolfsson et al. (2003) shows that online retailers have between 10 and 30 times more products than brick-and-mortar retailers. Aguiar and Waldfogel (2018) notes that the number of new music products between 2000 and 2008 tripled. Neither of these papers can separate changes in search costs from other changes in technology such as production costs. My findings suggest that discovery cost reductions were likely responsible for much of online product proliferation.

Estimates in Columns (3) and (4) show that average game entrant ratings fell after re-categorization relative to non-game entrant ratings. The point estimate at the game/non-game level is negative and not statistically significant, but the estimate at the app-type level is significant at the 99% confidence level. It shows that average ratings fell by 0.061 stars. This estimate is small but economically meaningful. It represents a 1.5% drop compared to the average app rating (4.0 stars), and a nearly 2% drop compared to the average game app rating (3.4 stars). This suggests that reductions in discovery costs reduce entrant quality.

Section 4.1 shows that discovery costs fell through two main channels, an increase in category informativeness and a reduction in congestion as measured by the number of apps per category. Some app-types were more affected than others. If supply-side entry and quality effects are driven by discovery cost changes, they should display similar heterogeneity. For example, consumers became more informed about app-types that did not feature in the pre-policy categories. Entry in these app-types should increase more as well. I test for such heterogeneity in Table 6. As in panel (a) of Table 4, Columns (1) and (2) test for the effects of changes in the informativeness of categories by comparing app-types with and without pre-existing categories. As in panel (b) of Table 4, Columns (3) and (4) test for the effects of changes in congestion costs by comparing small and large app-types among those split from two pre-policy game categories. Small app-types had fewer apps before re-categorization and they experience greater decreases in congestion after re-categorization.

Estimates confirm that average effects in Table 5 were primarily driven by app-types with greater changes in discovery costs. Entry increased more for game app-types that did not have pre-existing categories as compared to game app-types with pre-existing categories (relative to non-game types). Entrant quality also fell by more for these app-types. Among the app-types split off from pre-policy game categories, smaller app-types also had bigger changes in entry and entrant quality. App-types whose discovery costs were less affected by the policy have statistically null changes for both of the main supply-side outcomes.[57] These results are also robust to alternative outcomes such as the absolute number of entrants and to alternative measures of quality (Table C9).

Results in Tables 5 and 6 suggest that changes in consumer discovery costs are the main driving mechanism for product assortment changes in this market. The heterogeneity in entry effects also reflects theoretical predictions about the effects of discovery cost changes for different product types. Bar-Isaac et al. (2012) predict that "niche" products that benefit more from search cost reductions will experience the greatest increase in assortment. In this setting, the definition of "niche" products can include either app-types that had no pre-existing categories or "small types" that were relatively marginalized under the initial category structure.

[57] I also show this by looking at app-type specific treatment effects in Table C8.

Table 6: Entry and Quality Difference in Differences Estimates: Discovery Cost Channels

Outcome Variable:	ln(N Entrants) (1)	Mean Entrant Rating (2)	ln(N Entrants) (3)	Mean Entrant Rating (4)
Games × Post	0.251*	-0.015	-0.086	0.078
	(0.137)	(0.032)	(0.069)	(0.076)
Games × Post × No Pre-Existing	0.550***	-0.082**		
	(0.138)	(0.036)		
Games × Post × Small Type			0.833***	-0.177**
			(0.202)	(0.086)
Unit of Observation	App-Type	App-Type	App-Type	App-Type
Time Period	Jan 12 / Dec 14	Jan 12 / Dec 14	Jan 12 / Dec 14	Jan 12 / Dec 14
Sample	All	All	All Non-Games + Action, Arcade Card and Casino	All Non-Games + Action, Arcade Card and Casino
Year/Month FE	•	•	•	•
App-Type FE	•	•	•	•
Observations	1,470	1,470	980	980
R-squared	0.976	0.676	0.959	0.746

Notes: The sample period in all columns is January 2012-December 2014. Data in all columns includes monthly observations at the app-type level. Sample in Columns (1) and (2) includes all game and non-game app-types. Sample in Columns (3) and (4) includes all non-game app-types and Arcade, Action, Card and Casino game types. Outcomes in Columns (1) and (3) are the natural log of the number of entrants in each app-type. Outcomes in Columns (2) and (4) are mean average ratings of new entrants in each app-type. Controls include year and month fixed effects, app-type fixed effects and app-type specific time trends. The variable "Games × Post" is a dummy variable equal to 1 for games (or game app-types for even columns) during and after March 2014. The variable "Games × Post × No Pre-Existing" is equal to 1 during and after March 2014 only for app-types that did not have pre-existing categories before March 2014. The variable "Games × Post × Small Type" is a dummy equal to 1 during and after March 2014 only for Action and Casino game app-types. Standard errors in all columns are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

Correlation between the two supply-side outcomes suggests after discovery costs fell, more infra-marginal lower quality apps found it profitable to enter into the market. In the Online Appendix, I test another possible explanation for the quality effects. Fishman and Levy (2015) propose a model where incentives to produce high quality products decline as search costs fall. This happens because consumers are more likely to look through a number of high quality products and choose the best match and downloads for each high quality product decline to the point where firms do not recoup their quality investment. With this mechanism, the entire distribution of entrant quality should shift down. In Online Appendix C.9 I test for changes at different moments of the entrant rating distribution rather than just the average. In Table C10 I show that ratings fall more at the lower end of the entrant rating distribution than at the top. This suggests additional entry of infra-marginal apps is the primary mechanism driving reductions in quality.

4.3 Robustness and Additional Results

I test for several possible concerns with the research design or the data. I summarize these robustness checks below and provide additional details in the Online Appendix.

- Timing Tests (Online Appendix C.8): Timing tests for the main supply-side outcomes (Figure C3 in the Online Appendix) shows that the treatment effects do not "activate" before the re-categorization announcement. Quality measure timing tests also suggest that ratings based quality proxies are objectively comparable before and after re-categorization. Additional discussion is in Online Appendix C.8. Timing tests for download outcomes are in Online Appendix C2 and in Figure C1.

- Changes in monetization technology and prices (Online Appendix C.10): Figure C4, showing the share of new paid game and non-game apps entering over time shows that there were no differential changes between games and non-games in app monetization technology. Table C11 shows there are no statistically significant changes in average paid app prices or in paid entrant prices after re-categorization between games and non-games. Table C12 also shows there were no changes in app-level game prices for apps with differential changes in the number of other apps in their category.

- Changes in advertising (Online Appendix C.11): One possible contemporaneous confounder is advertising by Google for Google Play Store games. I do not directly observe ad spending, but I use Google Trends as a general proxy for consumer awareness. Comparisons of Google Trends for "Android Games" against "Android Apps" and "iOS games" search terms in Figure C5 in the Online Appendix show that there are no changes happening after re-categorization. There is no evidence of other contemporaneous confounders, such as the release of new phones or changes in consumer demographics or app usage.[58]

- Developer switching (Online Appendix C.12): A concern with the entry results is that developers with fixed budgets switch between producing games and non-games. The estimated treatment effect would be an upper bound as the control

[58]Nielsen Insights and Mobile NetView surveys from 2012 to 2014 suggest there is no change in growth rates of per-consumer mobile game usage. There were also no changes in demographics of Android users.

group is affected by treatment. Figure C6 in the Online Appendix shows that non-game developers switch to game production only for two months after re-categorization. By contrast, period specific coefficient estimates (the timing tests in Figure C3) show that monthly treatment effects for entry and other outcomes of interest *do not* decline after switching ends.

5 Structural Model

Estimates in Sections 4.1 and 4.2 show that re-categorization decreased discovery costs, which increased entry and reduced entrant quality. Although immediate reductions in discovery costs and longer run product variety growth should benefit consumers, additional entry also created additional congestion which plays a role in this market (Column 7 of Table 4). The overall welfare implications of the re-categorization policy are unclear. Over 80% of apps in the market are free, so the benefits of easier discovery or additional entry are not easily measurable. There was also already a large number of products in the market before the increase in entry. Additional products may not add much to consumer welfare. The fall in new product quality could reduce consumer welfare relative to a counterfactual world without re-categorization with fewer higher quality entrants. This calls for a model-based approach to measure and decompose the various welfare effects.

A decomposition of product assortment related and discovery cost related welfare effects provides policy relevant evidence. Policy-makers are concerned that high discovery costs foreclose entry in online markets and hurt consumer welfare. It is important to know whether changes in entry in online markets are negligible for consumer welfare, or if they are countered by changes in market congestion or entrant quality. In this section I set up and estimate a demand model that allows me to measure changes in consumer welfare in the market. I then decompose the various welfare effects.[59]

5.1 Demand Model

Consumer i chooses to download app $j \in \{1, 2, .., N\}$ in market/period t. The utility she receives from downloading app j of type c (for the moment suppressing market

[59]I do this without a formal supply model by relying on reduced form evidence for counterfactual entry and quality patterns. In a previous versions of this paper, I introduced an incomplete information static entry model to compute counterfactuals and perform welfare analysis for a subsample of free apps. This produced qualitatively similar effects but required numerous restrictive assumptions.

t notation) is:

$$\begin{aligned} u_{ijc} &= \delta_{jc} + \zeta_{ic} + (1-\sigma)\epsilon_{ijc} \\ &= X_{jc}\beta + \xi_{jc} + \zeta_{ic} + (1-\sigma)\epsilon_{ijc} \end{aligned} \tag{5}$$

where X_{jc} are observable app and app-type characteristics and ξ_{jc} are unobservable product and app-type characteristics. An app's "type" is defined according to the classification described in Section 3.2. ϵ_{ijc} is a consumer/app specific demand shock with an iid EVT1 distribution (mean zero, standard deviation normalized to 1).[60] ζ_{ic} is a consumer/app-type specific demand shock, such that $[\zeta_{ic} + (1-\sigma)\epsilon_{ijc}]$ is also EVT1 distributed. The consumer/app-type specific shock allows for correlation in consumer preferences across apps within app-types, parametrized by σ. The model otherwise abstracts from unobservable consumer heterogeneity.[61] The consumer can also pick an outside option of not downloading anything and receive $u_{i0} = \epsilon_{i0}$.

Equation 5 describes a standard nested logit model. Such a model likely poorly captures consumer preferences in this and many other online markets with substantial discovery frictions.[62] Similar to Ackerberg and Rysman (2005), I introduce additional shifters - R_{jc} - that proxy these frictions and capture competition for unobservable store space that attract scarce consumer attention. R_{jc} transforms the standard nested logit market share expression as follows:[63]

$$s_{jc} = \frac{exp(\frac{R_{jc}+\delta_{jc}}{1-\sigma})}{\sum_{j'\in c} exp(\frac{R_{j'c}+\delta_{j'c}}{1-\sigma})} \frac{\left[\sum_{j'\in c} exp(\frac{R_{j'c}+\delta_{j'c}}{1-\sigma})\right]^{1-\sigma}}{\sum_{c'\in\{1,..,C\}} \left[\sum_{j''\in c'} exp(\frac{R_{j''c'}+\delta_{j''c'}}{1-\sigma})\right]^{1-\sigma}} \tag{6}$$

[60] Consumer preference for variety comes primarily through this shock. This is a reasonable assumption for a market with many minimally horizontally differentiated product versions such as "Angry Birds Space" and "Angry Birds Star Wars".

[61] There could be additional unobservable heterogeneity in preferences for quality. Estimating a random coefficients model using aggregate app data is computationally challenging with hundreds of thousands of products and a small number of markets/time periods. It may be possible to limit the choice set to apps that appear in the top 500 best-seller list at least once, as in Leyden (2018). This raises concerns about sample selection and there is likely not enough variation in market shares to identify the distribution of unobservable heterogeneity (Berry et al. 2004, Armstrong 2016).

[62] As a robustness check, I also estimate a standard nested logit model. Results from this model are in Table D2 in Online Appendix D. Estimated σ in this model is bigger than 1, suggesting the model does not fit the data.

[63] I consider R_{jc} to be distinct from δ_{jc} as the variables entering into R_{jc} should reflect changes to utility coming from changes in app discovery costs rather than consumption utility. In practice, this distinction is conceptual in the main text. In the search model in Online Appendix D.1 R_{jc} enters the model non-linearly and is distinct from consumption utility.

where j' is an app that belongs to app-type c and j'' is an app that belongs to app type c'.

There are many possible specifications for R. I choose a parametric specification that separates R_{jct} into three sets of variables that capture the main sources of consumer search frictions discussed in Sections 2.1 and 4.1:

$$R_{jct} = +\gamma^1_{c,Pre/Post} + \gamma^2 ln(N_{c^*t}) + \gamma^3 ln(q_{jt-1}) \tag{7}$$

$\gamma^1_{c,Pre/Post}$ are time-varying app-type fixed effects. Each app-type has two dummies, one that is active before and one that is active after re-categorization.[64] Differences in pre/post app-type fixed effects capture average changes in consumer utility from downloading a type c app after re-categorization as compared to before, after conditioning on other observable app characteristics. These fixed effects capture changes in the informativeness of the categorization structure. As discussed in Section 4.1, some app-types such as Music or Educational games did not have pre-existing category spaces before re-categorization. After re-categorization, these app-types became easier to find for consumers, increasing overall consumer utility from these products. For other app-types such as Racing or Sports games, there should be minimal changes in informativeness.[65]

N_{c^*t} is the number of apps in the *category* of app j at time t. This variable captures the congestion externalities of increasing the number of apps in a category on the consumption of app j by reducing the probability that app j lands in a space that is visible to consumers. As described in Section 2.1, consumers look for products on the app store by browsing through categories and primarily see category-specific bestsellers and featured products. With more products in a category, a given product is less likely to appear to consumers. I show that increases in the number of apps in a category reduces aggregate app-type and individual app-level downloads in Table 4 in Section 4.1.[66] c^* represents the actual observed categorization of apps in the

[64]Formally, $\gamma^1_{c,Pre/Post} = \sum_c \gamma^1_{c,Pre}(D_c \times D_{Pre}) + \sum_c \gamma^1_{c,Post}(D_c \times D_{Post})$, where D_c is an app-type c dummy, D_{pre} is a dummy equal to 1 before March 2014 and D_{post} is a dummy equal to 1 after March 2014.

[65]Although these fixed effects could capture other time-varying app-type specific demand shocks, Figure D1 in the Online Appendix shows that they are identified by changes in downloads and market shares precisely around the re-categorization period.

[66]N_{c^*t} could also affect consumer demand through changes in the number of substitutes rather than through congestion externalities, but reduced form results suggest otherwise. In Section 4.1 I show that the number of apps in a category matters for downloads even when the number of substitutes to an app is held constant. I find the same magnitude of effects using long term changes in Online Appendix C.4. In Online Appendix C.10 I also show that changes in the number of apps in a category does *not* affect paid app prices. If the number of apps in a category primarily affected

store rather than app-type (c), although after re-categorization c and c^* coincide for app j.

q_{jt-1} is the number of downloads of app j in period $t-1$. It captures intertemporal externalities in the mobile app market: the effects of past popularity on present demand for app j. This reflects the underlying design of the app store and many online markets. An app with more downloads in period $t-1$ is more likely to be on a best-seller list and highly visible in period t.[67] Apps are also frequently advertised on and off the store. Apps with more past downloads are likely to invest more in advertising.

There are many possible models of app store consumer search behaviour. For example, Online Appendix D.1 sets up a model of simultaneous consumer search and demand with heterogeneous consideration sets.[68] However, identifying a specific micro-founded search model is not be credible without detailed individual click-stream data. The demand model presented in this section accounts for fundamental sources of product discovery frictions in this and many other online markets. It is consistent with a broad set of implications from theoretical and empirical search literature and can be credibly estimated with app-level data.

5.2 Demand Estimation, Identification and Results

Inverting the market share of app j of app-type c at time t produces the following linear estimating equation:

$$ln(\frac{s_{jct}}{s_{0t}}) = X_{jt}\beta + \gamma^1_{c,Pre/Post} + \gamma^2 ln(N_{c^*t}) + \gamma^3 ln(q_{jt-1}) + \sigma ln(s_{j|c,t}) + \xi_{jct} \tag{8}$$

where s_{0t} is the market share of the outside option.[69] $s_{j|c,t}$ is the within-app-type market share in period t. I estimate the model by solving this equation for the

competition intensity, it should affect paid app prices.

[67]An alternative variable may be the actual past best-seller rank of an app. I choose not to use it for several reasons. Ranks change on a weekly basis, but my estimation is done on a monthly basis for computational reasons. Also, the implications of ranking as number 1 may be different for different categories. Downloads account for this.

[68]This model produces a similar market share structure to the linear model in the main text, with some more non-linearity in the specification of R_{jct}. Estimates from this model are in Table D2 and are similar to the linear estimates in Table 7.

[69]I assume that total market size is twice the maximum total number of purchases observed in a time period. The Android handset market is growing over time, so another possible assumption is to match total market size to the number of US Android handsets. Since I include year/month fixed effects other normalizations do not change parameter estimates qualitatively or quantitatively.

structural error term ξ_{jct} and interacting it with instruments to form the GMM criterion function. I include app-type, year/month, app-age and developer fixed effects to control for unobservable heterogeneity.

Changes in characteristics shift consumer utility and affect purchase decisions and observable market shares, identifying the β parameters. Changes in N_{c^*t} and q_{jt-1} similarly identify γ parameters. For example, if two apps with the same quality are located in categories that have a different number of apps, the app in the more populous category should be harder to find and its market share should be lower. If the number of apps in the category of app j is increasing, holding everything else constant, changes in market shares over time should also identify γ^2. Changes in average downloads for each app-type after re-categorization identifies γ^1 parameters.[70]

Prices and within-type market shares are likely correlated with unobservable product quality (ξ). Products with higher ξ have more demand, higher prices and higher within-type market shares, meaning that the price coefficient and σ are biased. Lagged app downloads have a similar problem, as products with higher unobservable quality are more popular over time.

I use characteristics-based instruments to address this endogeneity: average ratings, size in MB, and number of screenshots of other type c apps, excluding app j. Intuitively, higher average quality for other products is going to change app j's prices through competition independently of j's own demand shock. A concern with these instruments are endogenous entry decisions, which could make product assortment and average app-type quality vary in response to demand shocks. I assume that apps make their entry decisions before learning the realization of consumer demand shocks in period t. Consumer demand in large online markets is uncertain and there are many surprise successes and failures.[71] After controlling for app-type and time specific demand shocks using app-type and year/month fixed effects, entry variation should be driven by supply-side fixed/entry cost shocks.

There may still be concerns about correlation between the number of apps in a category (N_{c^*t}) and category/time varying unobservable demand shocks. To address this I instrument for N_{c^*t}. My instrument is the residual of a regression of the number of apps in category c^* in period t on the number of apps in category c^* in period $t-1$, controlling for total category downloads in period t and mean category rating in period t. The intuition of this IV is that realized period t total downloads and

[70]Figure D1 in the Online Appendix shows that γ_1 are not identified through variation other than the re-categorization shock. This is analogous to the reduced form estimates shown in Section 4.1.

[71]An example of a surprise success is an app matching users' faces to paintings (ArtNet.com). For an opposite example, the app Sip by the creators of the popular site ProductHunt failed to make an impact despite substantial investment and advertising (ProductHunt blog).

category ratings control for any demand side-shocks that influence entry decisions. The remaining variation in the number of apps between period $t-1$ and period t is driven by supply side cost shocks.[72]

I use $\Delta q_{jt-2} = ln(q_{jt-2}) - ln(q_{jt-3})$ as an instrument for lagged downloads. This instrument exploits over-time correlation in downloads for a given app (Blundell and Bond 1998). An app's downloads in period $t-2$ and $t-3$ share a history of demand shocks with its downloads in period $t-1$ (e.g., both were affected by the app being featured on the store in some previous period). But downloads in period $t-2$ and $t-3$ are not affected by the demand shock in period $t-1$. Differencing removes persistent unobservable heterogeneity, making this a valid IV.[73]

I estimate the model using data on all game apps from March 2012 to December 2014 (also excluding March 2014). Results are in Table 7. As mentioned above, I include developer fixed effects in all specifications and additional controls such as app age fixed effects. Standard errors are clustered at the app level.

Estimates of σ, β and γ coefficients are in Table 7. Estimated coefficients for the number of apps in a category and the lag of past downloads confirm anecdotal evidence from Section 2.1 and reduced form evidence from Section 4.1. They suggest that there are product discovery frictions and congestion costs in the mobile app market. The coefficient on the number of apps in the category is negative and statistically significant. It suggests that an increase in the number of apps in app j's category reduces demand and consumer utility.

The coefficient on lagged downloads is positive. It suggests that apps with more past downloads are easier to find by consumers. It is also consistent with previous findings in the empirical literature on online product ranking-based discovery frictions (e.g., Ursu 2018).

Figure 3 plots estimates of app-type specific differences in pre/post re-categorization fixed effects. It shows substantial heterogeneity across app-types. On average, the ten app-types that did not have a pre-existing category (Adventure, Board, Educational, Family, Music, Role Playing, Simulation, Strategy, Trivia and Word) experience larger average increases in utility as compared to the eight app-types that had explicit categories before (Action, Arcade, Card, Casino, Casual, Puzzle, Racing and Sports).[74] These effects are quantitatively large: on average, utility increases by

[72] Estimates of this supporting regression are in Table D1 in Online Appendix D.

[73] It is also possible to take first differences of Equation 8 and use past lags of product j's downloads as instruments. This approach takes advantage of Arellano and Bond (1991)'s moment conditions, and has been previously used by Sweeting (2013) and Aguirregabiria and Ho (2012). However, in my application first differences remove much of the variation in the data at the category level and produce weak IVs.

[74] The main exception for this group is Action games, which have a change in fixed effects com-

Table 7: **Demand Model Parameter Estimates**

γ Estimates	
ln(N Apps in Category)	-0.393***
	(0.026)
ln(Lag App Downloads)	0.034***
	(0.004)
σ and β Estimates	
σ	0.709***
	(0.027)
Price	-0.836***
	(0.111)
ln(Size in MB)	0.046***
	(0.006)
N Screenshots	0.008***
	(0.001)
Video Preview Dummy	0.072***
	(0.017)
Paid App Dummy	0.107
	(0.194)
App Age FE	•
App Rating FE	•
Year/Month FE	•
App-Type FE	•
Developer FE	•
Observations	4,152,147
R-squared	0.390

Notes: Sample includes game app-month observations from March 2012 to December 2014 (excluding March 2014) in the Google Play Store. "App Rating FE" are a set of dummies representing the average rating of app j in period t within 0.5 stars. Apps with 2 stars or less are the baseline group for the "App Rating FE." Instruments for price and for σ include the average ratings of other apps in the same app-type, the average number of screenshots of other apps in the category, and the average size of other apps in the category. Instruments for lagged downloads of app j include differences in the lags of app j downloads (2 and 3 periods before period t). The instrument for the number of apps in the category is described in the main text and is the residual of the regression in Table D1. Standard errors are clustered at the app level. *** p<0.01, ** p<0.05, * p<0.1.

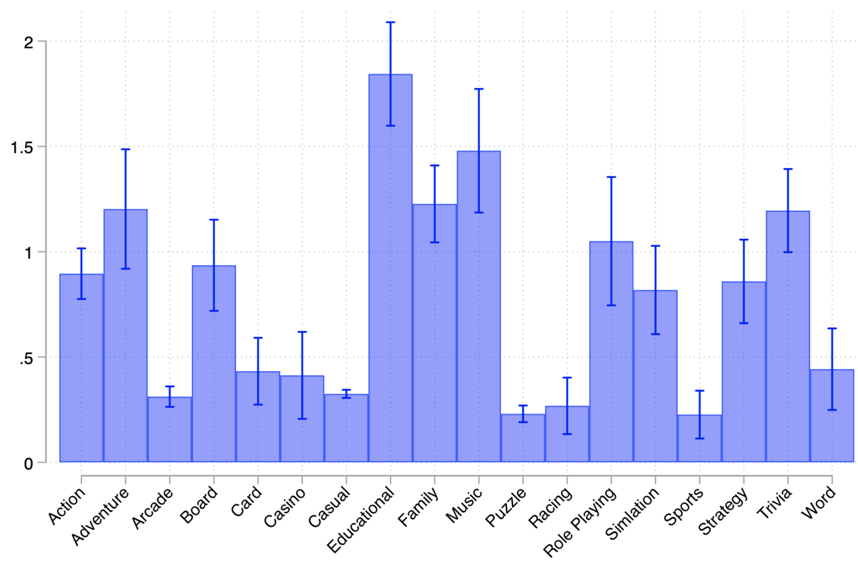
1 dollar for consumers from buying an app belonging to an app-type that did not have a category before the change, holding everything else constant.[75] This is con-

parable to some of the other app-types. This is possibly because it was grouped together with the Arcade app-type before re-categorization, leading to substantial improvement in the consumer search process.

[75]Re-categorization also increases average utility for consumers from purchasing other app-types,

sistent with reduced form evidence from Section 4.1, showing that re-categorization increased downloads for those app-types more than for the second group of app-types, and suggests that informativeness of the category structure increased after re-categorization.

Figure 3: Difference in App-Type Fixed Effects



Notes: Each column shows the difference in estimated app-type fixed effects based on the model described in Section 5.1: $\hat{\gamma}^1_{c,Post} - \hat{\gamma}^1_{c,Pre}$. 95% calculated confidence interval for this difference is shown.

Price coefficients show that demand for paid mobile games is relatively elastic. Median price elasticity for paid games is over 2. Median demand elasticity with respect to app size is approximately 0.25. Quality proxies (average rating, number of screenshots, presence of video preview) shift consumer utility in expected directions.

I tested several alternative specifications of the model for robustness. Results are in Table D2 in Online Appendix D. For example, I allow new apps to have different discovery frictions than incumbent apps. There are no statistically significant

although effects there are less than half the size on average (except for Action games, see previous footnote). This is likely because the informativeness changed for the other app-types as well, albeit in relatively minor ways. For example, a consumer looking for a Card game may be less confused about what kinds of Card games are in the "Cards" category (i.e., no family card games, no music card games).

differences in γ between the two types of apps. I also estimate a non-linear GMM model based on a micro-founded search and demand model developed in Appendix D.1. Results are qualitatively similar to those in the main text with respect to the main parameters of interest.

5.3 Consumer Welfare Analysis and Decomposition

In this section I calculate the overall consumer surplus change from re-categorization and decompose it. The decomposition separates surplus from changes in category informativeness; surplus from the immediate changes in congestion costs after re-categorization (holding fixed the number and composition of apps in the market); surplus from additional entry after re-categorization; surplus from changing congestion due to increased entry; and surplus from the changing quality of new apps entering the market relative to a counterfactual state with higher quality (but the same total number of) entrants.[76] I focus on consumer surplus as this is the most important metric for policy-makers in digital markets.

App j of type c is located in category c^* before re-categorization. After re-categorization it is located in category c, which coincides with its type. The number of apps in category c^* is N_{c^*}. The number of apps of type c is N_c. A consumer who downloads app j of type c receives baseline utility $\hat{\delta}_j = X_j\hat{\beta}$. That consumer also receives/pays "discovery costs," $\hat{R}_j(\hat{\gamma}^1_{c,Pre/Post}, N_{c/c^*})$, which is linearly additive to $\hat{\delta}_j$ and defined in Equation 7 as a function of the characteristics associated with the difficulty of discovering this app. The relevant components of $\hat{R}_j(.)$ are $ln(N_{c/c^*})$, the number of other apps in the category of app j, and $\hat{\gamma}^1_{c,Pre/Post}$, pre and post re-categorization app-type specific fixed effects capturing the informativeness of the categorization structure for finding apps of type c. Lagged app downloads also matter for discovery and I include them in the model, but I omit them from the exposition for the sake of clarity.[77]

Changes in consumer surplus are the difference between expected utilities under different counterfactual market configurations (number and type of apps in the market), converted into dollar values using the price coefficient from Table 7.[78] With the

[76] Other possible counterfactuals in this market relate to the optimal number of categories and optimal entry. Due to the specific nature of platform re-designs, these are impossible to capture without observing other re-categorization events. There are no such events during my sample period.

[77] See the last paragraph of this section for additional details.

[78] There may be a concern that the price coefficient, estimated from paid apps, cannot be used to calculate free apps' dollar value. However, an interpretation of the price coefficient in additive random utility models is of the marginal value of money, or the marginal value of income for

demand model from Section 5.1, expected consumer utility is:

$$EU\left(\Omega, R(\gamma^1, N)\right) = ln\left(1 + \sum_{c\in\{1,..,C\}}\left[\sum_{j\in\Omega_c} exp\left(\frac{\hat{\delta}_j + \hat{R}_j(\hat{\gamma}^1_{c,Pre/Post}, N_{c^*/c})}{1-\sigma}\right)\right]^{1-\sigma}\right) \tag{9}$$

where Ω_c is the set of apps of type c and $\Omega = \{\Omega_1, \Omega_2, ..., \Omega_c, ..., \Omega_C\}$.

I refer to different levels of EU below as functions of the variables that change in the counterfactuals, Ω, γ^1 and N. Expected utility is calculated as in Equation 9 throughout and to avoid repetition I do not re-write the expression each time. For a given app j, $\hat{\delta}_j$ does not vary in any of the counterfactuals. The number and type of apps in the market, Ω, does vary, which also changes $\hat{R}_j$. I describe computing immediate effects from re-categorization and longer-run effects that incorporate entry changes below.

5.3.1 Immediate Welfare Change from Re-Categorization

I first calculate expected consumer utility just before re-categorization. I use February 2014 as the baseline period and define product assortment in that period as Ω^{Pre}. Apps are separated into six categories, with $N_{c^*,Pre}$ apps in each category c^*. The fixed effects associated with each app-type c are $\hat{\gamma}^1_{c,Pre}$. Expected utility for this configuration of products is $EU_1 = EU\left(\Omega^{Pre}, \hat{R}(\hat{\gamma}^1_{Pre}, N^*_{Pre})\right)$. N^*_{Pre} summarizes the number of apps in each of the six pre-recategorization categories.

Next I calculate expected utility with the same product assortment but with a larger number of categories. I know the app-type of each app j in February 2014 and substitute N_{c^*} with N_c when calculating $\hat{R}_{jc}$ and $EU_2 = EU\left(\Omega^{Pre}, \hat{R}(\hat{\gamma}^1_{Pre}, N_{Pre})\right)$. N_{Pre} denotes the number of apps in each of the eighteen post-recategorization categories/ app-types.[79] EU_2 maintains the *original* informativeness level, representing a counterfactual where the additional categories are not necessarily informative

consumers in the market (McFadden 1981, Anderson, De Palma and Thisse 1992). Free and paid apps fundamentally exist in the same market and in the same characteristics space. I assume that average differences between the two are captured by the "paid app" fixed effect, which enters into the expected utility consumers receive from app j ($\hat{\delta}_j$). While there are costs to free games that vary across games, I assume that these are captured using other app-level quality proxies, such as an app's average rating. An analogy could be made to the older transportation literature, where the price of one mode of transportation (e.g., bus fare) helps calculate the value of other modes of transportation (e.g., cycling).

[79] This is calculated based on the app-type of each app in the February 2014 choice set.

about their contents.[80] Next I calculate expected utility with the same choice set and number of apps, but the new category informativeness level ($\hat{\gamma}^1_{Post}$): $EU_3 = EU\left(\Omega^{Pre}, \hat{R}(\hat{\gamma}^1_{Post}, N_{Pre})\right)$.

$\frac{EU_2 - EU_1}{-\beta_{price}}$ measures the per-consumer per-month effect of changing congestion frictions but holding informativeness and overall choice set constant. $\frac{EU_3 - EU_2}{-\beta_{price}}$ is the per-consumer effect of changing the informativeness of the new categorization structure. Together the two differences in expected utility add up to the immediate effects of re-categorization through changes in discovery costs, holding the choice set constant.

5.3.2 Longer Run Welfare Change from Re-Categorization

Next I compute the longer-run welfare effects of re-categorization that incorporate changes in entry. I use December 2014, the last period available in my sample to do so. Expected utility in the factual market is $EU_4 = EU\left(\Omega^{Post}, \hat{R}(\hat{\gamma}^1_{Post}, N_{Post})\right)$, where Ω^{Post} is *actual* product assortment in December 2014 and $N_{c,Post}$ is the *actual* number of type c apps in December 2014. $\frac{EU_4 - EU_3}{-\beta_{price}}$ measures the overall welfare difference between February and December 2014, but this difference does not capture the effect of the policy. The relevant welfare difference needs to compare factual welfare in December 2014 to a *counterfactual December 2014 market without re-categorization*.

Section 4.2 shows that both the total number of entrants and the composition of entry (e.g., quality) was affected by re-categorization. Without re-categorization, there would have been fewer entrants in December 2014 but average entrant quality would have been higher relative to observed entry. I separately isolate these using two counterfactual December 2014 markets. The first counterfactual market, denoted as $Post'$, has fewer entrants than the factual December 2014 market but the same average entrant quality. I define the choice set in this counterfactual market as $\Omega^{Post'}$ and the number of type c apps in this market is $N_{c,Post'} \leq N_{c,Post}$.[81] The second counterfactual market, denoted as $Post''$, represents a December 2014 choice set where *both* the number and quality of entrants is different compared to the factual December 2014 market. To isolate the effect of quality the number of type c apps in the second counterfactual market is $N_{c,Post'}$, the same as in the first counterfactual market. However, the set of apps is different: $\Omega^{Post''} \neq \Omega^{Post'}$. The difference

[80] It is possible to imagine categories with uninformative titles such as "Card & Casino 1" and "Card & Casino 2."

[81] I proxy quality using ratings such that $\frac{\sum_{j \in \Omega_c^{Post'}} \text{Avg. Rating}_j}{N_{c,Post'}} = \frac{\sum_{j \in \Omega_c^{Post}} \text{Avg. Rating}_j}{N_{c,Post}}$ for each app-type c.

between $\Omega^{Post''}$ and $\Omega^{Post'}$ is average app quality, proxied by ratings, which is set at the February 2014 level and higher than in the $Post$ or $Post'$ markets.[82]

Computing $\Omega_c^{Post'}$, $N_{c,Post'}$, $\Omega_c^{Post''}$ and $N_{c,Post''}$ is an empirical challenge with two solutions. One solution involves setting up and estimating a supply model, such as a dynamic incomplete information app entry model. With estimates from this model, it is possible to compute equilibrium demand and supply under different store categorizations. I do not take this approach as it would require me to impose numerous strong assumptions.[83] Estimating an entry model with many players also introduces severe computational challenges.

Instead, I calculate the choice set in state $Post'$ by randomly removing apps that entered between March 2014 and December 2014 from the December 2014 choice set. Reduced form estimates from Section 4.1 tell me how many entrants to remove. I remove entrants to match both the aggregate entry treatment effects from Column (1) of Table 5 and app-type specific entry treatment effects from Table C8 from the Online Appendix.[84] I repeat this exercise 100 times to ensure randomization is not driving results.[85] Since I randomly omit entrants, average entrant quality remains the same in state $Post'$ as compared to the factual market in December 2014.

I similarly compute the choice set for counterfactual state $Post''$. I remove the same number of March/December 2014 entrants from the factual December 2014 choice set as in state $Post'$, but I remove relatively more entrants from the lower end of the quality distribution in each app-type. I do this so that in addition to matching entry effects from Tables 5 and C8, I also match app-type specific quality treatment effects from Table C8. For example, reduced form estimates show that re-categorization increases entry for Family games by 65% and reduces the average quality of entrants by 0.1 stars. I remove 39% of March/December 2014 Family game entrants ($\frac{1}{1+0.65}$), oversampling low rating entrants such that the average rating of the remaining entrants is 0.1 stars lower than the average rating of the full set of entrants in December 2014. I also repeat this exercise 100 times to ensure that the

[82] Quality is such that $\frac{\sum_{j\in\Omega_c^{Post''}} \text{Avg. Rating}_j}{N_{c,Post'}} \geq \frac{\sum_{j\in\Omega_c^{Post}} \text{Avg. Rating}_j}{N_{c,Post}}$ for each app-type c.

[83] For example, I would have to introduce assumptions about the information structure of entrants in the market, fixed and entry costs, and product design and location decisions by entrants.

[84] In aggregate, 146,688 game apps entered between March 2014 and December 2014. Based on Column (1) of Table 5, factual entry is 34% higher than without re-categorization. This means that $\frac{146{,}688}{1+0.34} = 109,469$ game apps would have entered the market without re-categorization, and I need to remove $146,688 - 109,469 = 37,219$ entrants. I choose which game apps to remove by similarly matching app-type specific entry rates in Table C8.

[85] Since I remove entrants randomly, there may be very popular apps (i.e., Angry Birds) that disappear under some draws.

initial randomization is not driving the results.[86]

After determining the set of December 2014 apps in the two counterfactual states, I calculate three additional expected utilities. $EU_5 = EU\left(\Omega^{Post'}, \hat{R}(\hat{\gamma}^1_{Post}, N_{Post'})\right)$ measures expected utility for the average consumer in state $Post'$, where entry is lower than in factual December 2014. The second expected utility is $EU_6 = EU\left(\Omega^{Post''}, \hat{R}(\hat{\gamma}^1_{Post}, N_{Post'})\right)$. It measures expected utility for the average consumer in state $Post''$, where entry is lower than in factual December 2014 but entrant quality is higher. To isolate the effects of changing congestion, I calculate $EU_7 = EU\left(\Omega^{Post}, \hat{R}(\hat{\gamma}^1_{Post}, N_{Post'})\right)$. This is the expected utility consumers receive when entry is at the *factual* December 2014 level, but congestion costs are at the *counterfactual* $Post'$ lower entry level.

The "gross" long-run welfare effect of re-categorization if only the number but not the average quality of entrants or congestion changes is $\frac{EU_7-EU_5}{-\beta_{price}}$. Since the number of type c apps is equal in states $Post'$ and $Post''$, the difference $\frac{EU_5-EU_6}{-\beta_{price}}$ isolates the long-run welfare effects of re-categorization through only changes in entrant quality. The difference $\frac{EU_4-EU_7}{-\beta_{price}}$ isolates the welfare effect of increasing long-run congestion costs as a result of the increase in variety due to re-categorization. Summing up these three effects gives the net overall net effect of re-categorization on consumer welfare through changes in product assortment. It is equal to $\frac{EU_4-EU_6}{-\beta_{price}}$.

In the equations and discussion above I omitted lagged app downloads (log of q_{jt-1} in Equation 7) for the sake of clarity, but they enter into $R(.)$ throughout. Lagged app downloads with a restricted choice sets in states $Post'$ or $Post''$ should not be equal to factual lagged app downloads in factual state $Post$. I recompute lagged app downloads in each state to account for this.[87]

[86] Randomization seeds are consistent across $Post'$ and $Post''$ replications.

[87] Lagged app downloads in period t depend on market shares in period $t-1$, which in turn depend on the market shares in period $t-2$ and so on. The full set of paths of market shares for each simulation is essentially impossible to recover under my randomization approach. In every period, I would allow the entry of an additional restricted set relative to the factual set. I would have to do multiple repetitions for each period to ensure that randomization is not driving the results. This would very quickly "blow-up" the number of possible paths. Instead, I assume that the choice set in the last period is the same as the choice set in the previous period and re-compute equilibrium market shares using a contraction mapping. I do this for both "factual" and "counterfactual" states to ensure consistency. Since the "lagged choice set" with this approach is larger than the true lagged choice set, this understates lagged app downloads and welfare effects coming from this source of discovery frictions.

5.3.3 Estimates

Table 8 shows the decomposition of consumer surplus changes. I show mean entry, congestion and quality effects based on the randomized procedure outlined above.[88]

Table 8: **Estimated Changes to Monthly Per-Consumer Surplus ($)**

	Immediate Effects	
1) Information Effect	$EU_3 - EU_2$	0.016
2) Congestion Effect	$EU_2 - EU_1$	0.025
	Long-Run Effects	
3) Mean Gross Entry Effect	$EU_7 - EU_5$	0.087
4) Mean Entry Congestion Effect	$EU_4 - EU_7$	-0.037
5) Mean Entrant Quality Effect	$EU_5 - EU_6$	-0.006
6) Total Effects	$EU_3 - EU_1 + EU_4 - EU_6$	0.085

Notes: This table shows estimates of average per-consumer welfare effects using demand model estimates from Table 7. Gross Entry Effects, Entry Congestion Effects and Entrant Quality Effects are averages from 100 simulations. The full distributions of simulation effects are in Figure D2 in the Online Appendix.

Re-categorization has a total positive effect on consumer surplus, increasing it by 8.5 cents per-consumer per-month. In 12 months the average consumer gains approximately $1 in additional surplus relative to a counterfactual world where re-categorization does not happen. This is equivalent to 50% of the mean paid game's price. Across all Google Play Store consumers, annual welfare gains add up to $100 million as there are over 100 million Android users in the US in 2014.

The first two rows of the table show immediate consumer welfare gains coming from the re-categorization while *holding the choice set fixed*. These gains come from two sources: increased informativeness of the categorization structure, and changes in congestion because of re-categorization. The second change of 2.5 cents per consumer per month is larger than the first, 1.6 cents per consumer per month. Together they constitute nearly 50% of the total gains. Adding these up across all Android consumers in the US, they add up to nearly $50 million per-year. These estimates confirm previous results suggesting large consumer discovery frictions in online markets and suggest discovery frictions meaningfully respond to platform re-design.

[88] Figure D2 in the Online Appendix shows the full distributions of these effects under different random draws. The variation in effect size across different randomization seeds is quantitatively small.

The third row of Table 8 shows *gross* consumer welfare gains due to additional product entry in response to the re-categorization relative to a counterfactual world where re-categorization does not happen. Gross gains are approximately *twice* as big as immediate gains coming from lower discovery costs. Adding rows four and five gives the *net* welfare gains due to additional product entry. I find that consumers lose around 50% of gross variety welfare gains.

Consumers lose 40% of gross variety welfare gains because of increasing congestion costs (fourth row). In absolute terms, losses from additional congestion are 3.7 cents per-consumer per-month relative to a counterfactual world where re-categorization does not happen and entry does not increase. The costs add up to approximately $44 million per-year for all Google Play Store users. These estimates show that congestion externality costs in online markets are substantial and consumers do not fully benefit from additional product variety in the market.

Reduced form results raise questions about the welfare effects of changes in product quality. The counterfactual simulations suggest that changing the quality distribution of entering products affects consumer welfare less than increasing congestion but it is still economically significant. Reductions in the quality of new products relative to a world where re-categorization does not happen take away 7% of gross variety welfare gains, adding up to over $7 million per-year across all Google Play Store consumers. The magnitude of quality welfare effects is notable since at the time of re-categorization there is already a substantial number of apps in the market at all quality levels.

6 Conclusion

This paper examines the effects of changes in consumer discovery costs on competition and consumer welfare in online markets using new data from the Google Play mobile app store. I take advantage of a re-design of the mobile game section of the store. Reduced form estimates using changes in downloads show that the re-design improved consumer discovery by increasing the number of categories, which reduces the number of apps-per-category and congestion, and by making category titles more informative. Reduced form estimates show that the reducing in discovery costs increases game app entry but reduces average entrant quality. The supply-side results are the first causal empirical evidence on the effects of consumer discovery costs on non-price market outcomes. Product assortment is the key competitive outcome of concern for regulators of online markets because prices in these markets are often zero or uniform.

I set up a structural demand model estimating consumer utility parameters and parameters proxying discovery costs. This allows me to measure the overall welfare effects of the re-categorization and decompose them. It is important for policy-makers to understand the magnitude of congestion externalities in online markets and how these compare to consumer surplus gains from variety, as well as the direct welfare effects of lower discovery costs and changing product quality. I find per-consumer gains from the re-categorization are approximately \$0.085 per-month. This adds up to nearly \$50 millions additional welfare each year across all 100 million Google Play Store consumers in the US. A decomposition shows that increasing congestion and the changing quality of new products take away 50% of gross variety gains.

Policy makers are concerned with high consumer discovery costs in a number of online markets (WSJ.com). The main claim is that *high* discovery costs online foreclose potential entrants and reduce variety. This paper is the first to provide evidence of welfare implications for changing discovery costs online. My findings show that welfare gains from variety are counteracted by increasing congestion externalities and lower entrant quality. In my application welfare changes coming from congestion and product quality are not enough to overwhelm the variety effect, and consumers benefit.[89] However, welfare changes from increased congestion and from reduced entrant quality are still economically significant and annually add up to tens of millions of dollars of lost consumer surplus.

A \$2.4 billion EU fine to Google provides an interesting case (TheVerge.com). Google's policy in the search engine market (e.g., hotels, flights) was judged to have increased consumer search costs. The harm was not limited to existing competitors (e.g., Expedia). There was also potential consumer harm due to the foreclosure of future competitors.[90] Other online platforms face similar concerns. Spotify has been accused of promoting artists to whom it can pay lower royalty fees over higher quality but more expensive competitors (NY Times.com). Examining platform incentives in a setting where discovery costs affect equilibrium entry and product quality is an extension that should be pursued in future research.

[89] The platform also benefits, as evidenced by subsequent additional game and non-game categories Google introduced after my sample period.

[90] There are potential benefits to Google's search engine policies that my model does not capture. For example, the quality of Google's products is arguably higher than their competitors' qualities. Google may minimize the total search costs consumers pay.

References

Ackerberg, D. A. and M. Rysman (2005). Unobserved product differentiation in discrete-choice models: estimating price elasticities and welfare effects. *RAND Journal of Economics 36*(4), 771–789.

Aguiar, L. and J. Waldfogel (2018). Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music. *Journal of Political Economy 126*(2), 492–524.

Aguirregabiria, V. and C.-Y. Ho (2012). A dynamic oligopoly game of the us airline industry: Estimation and policy experiments. *Journal of Econometrics 168*(1), 156–173.

Anderson, S. P. and R. Renault (1999). Pricing, product diversity, and search costs: a bertrand-chamberlin-diamond model. *RAND Journal of Economics 30*(4), 719–735.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies 58*(2), 277–297.

Armstrong, T. B. (2016). Large market asymptotics for differentiated product demand estimators with economic models of supply. *Econometrica 84*(5), 1961–1980.

Bar-Isaac, H., G. Caruana, and V. Cuñat Martinez (2012). Search, design and market structure. *American Economic Review 102*(2), 1140–1160.

Berry, S., O. B. Linton, and A. Pakes (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies 71*(3), 613–654.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics 87*(1), 115–143.

Brynjolfsson, E., Y. Hu, and D. Simester (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science 57*(8), 1373–1386.

Brynjolfsson, E., Y. Hu, and M. D. Smith (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science 49*(11), 1580–1596.

Chen, Y. and T. Zhang (2017). Entry and welfare in search markets. *The Economic Journal 128*(608), 55–80.

Chevalier, J. A. and D. Mayzlin (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research 43*(3), 345–354.

Coey, D., B. J. Larsen, and B. C. Platt (2020). Discounts and deadlines in consumer search. *The American Economic Review (forthcoming)*.

Diamond, P. A. (1971). A model of price adjustment. *Journal of Economic Theory 3*(2), 156–168.

Dinerstein, M., L. Einav, J. Levin, and N. Sundaresan (2018). Consumer price search and platform design in internet commerce. *American Economic Review 108*(7), 1820–59.

Fershtman, C., A. Fishman, and J. Zhou (2018). Search and categorization. *International Journal of Industrial Organization 57*, 225–254.

Fishman, A. and N. Levy (2015). Search costs and investment in quality. *The Journal of Industrial Economics 63*(4), 625–641.

Fradkin, A. (2017). Search, matching, and the role of digital marketplace design in enabling trade: Evidence from airbnb. *Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb (March 21, 2017)*.

Gandhi, A., Z. Lu, and X. Shi (2014). Demand estimation with scanner data: Revisiting the loss-leader hypothesis. mimeo.

Goldmanis, M., A. Hortaçsu, C. Syverson, and Ö. Emre (2010). E-commerce and the market structure of retail industries*. *The Economic Journal 120*(545), 651–682.

Kesler, R., M. Kummer, and P. Schulte (2020). Competition and privacy in online markets: Evidence from the mobile app industry. In *Academy of Management Proceedings*, Volume 2020, pp. 20978. Academy of Management Briarcliff Manor, NY 10510.

Leyden, B. T. (2018). There's an app (update) for that. *mimeo*.

Liebowitz, S. J. and A. Zentner (2020). The challenges of using ranks to estimate sales. *Available at SSRN 3543827*.

Liu, Y., D. Nekipelov, and M. Park (2014). Timely versus quality innovation: The case of mobile applications on itunes and google play. *NBER Working Paper*.

Moraga-González, J. L. and Y. Sun (2020). Product quality and consumer search.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy 69*(3), 213–225.

Sweeting, A. (2013). Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry. *Econometrica 81*(5), 1763–1803.

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science 37*(4), 530–552.

Varian, H. R. (2007). Position auctions. *international Journal of industrial Organization 25*(6), 1163–1178.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, 641–654.

Zentner, A., M. Smith, and C. Kaya (2013). How video rental patterns change as consumers move online. *Management Science 59*(11), 2622–2634.

APPENDIX TO: "Consumer Product Discovery Costs, Entry, Quality and Congestion in Online Markets" (for Online Publication)

A Appendix A

A.1 List of Non-Game Categories

Table A1: Google Play Non-Game Categories

Books & Reference	Libraries & Demo	Productivity
Business	Lifestyle	Shopping
Comics	Media & Video	Social
Communications	Medical	Sports
Education	Music & Audio	Tools
Entertainment	News & Magazines	Transportation
Finance	Personalization	Travel & Local
Health & Fitness	Photography	Weather

A.2 Examples of Consumer Responses to Re-Categorization

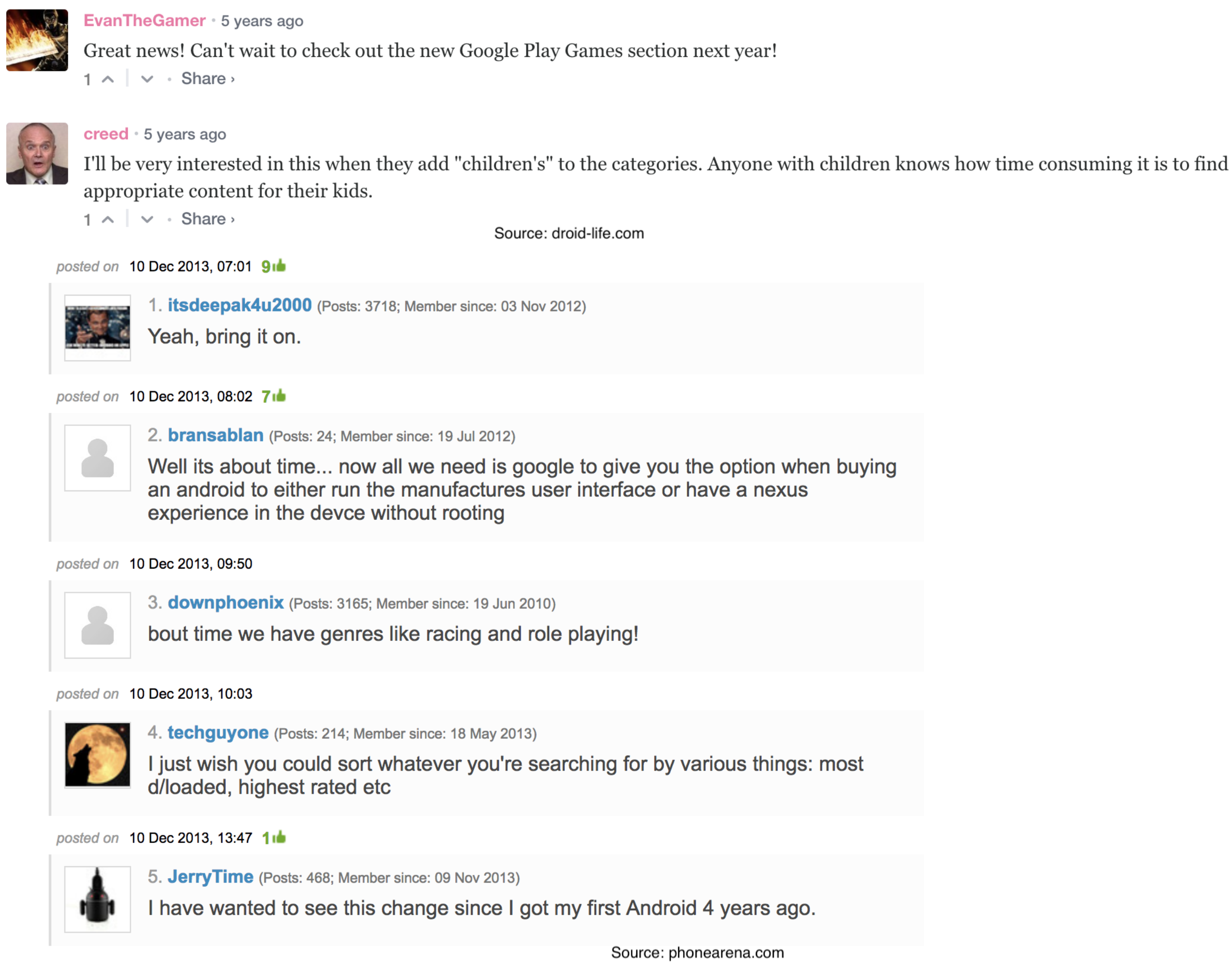


Figure A1

A.3 Consumer App Discovery Surveys

Figure A2 shows results from two consumer surveys taken during my sample period by Forrester and Nielsen. Both surveys asked app consumers how they discovered new products. 58% of Android consumers discover new products through "General browsing in the app store", according to Forrester, and 25% discover new products through more targeted browsing - looking at “top rated” or “most popular” app lists in the app store. Only a small share of consumers discover new apps through an internet search engine. Answers are similar in the Nielsen survey, with 40% of consumers browsing the app store to discover new products and only 27% using search engines.

Figure A2: Google Play Consumer Product Discovery Surveys

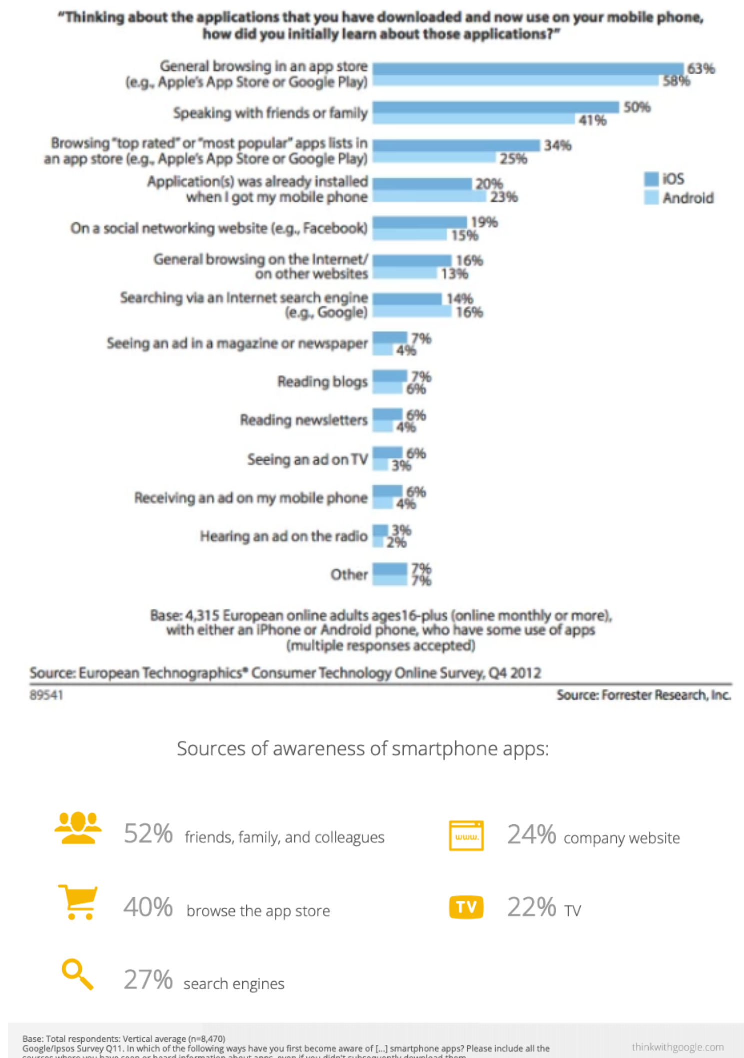
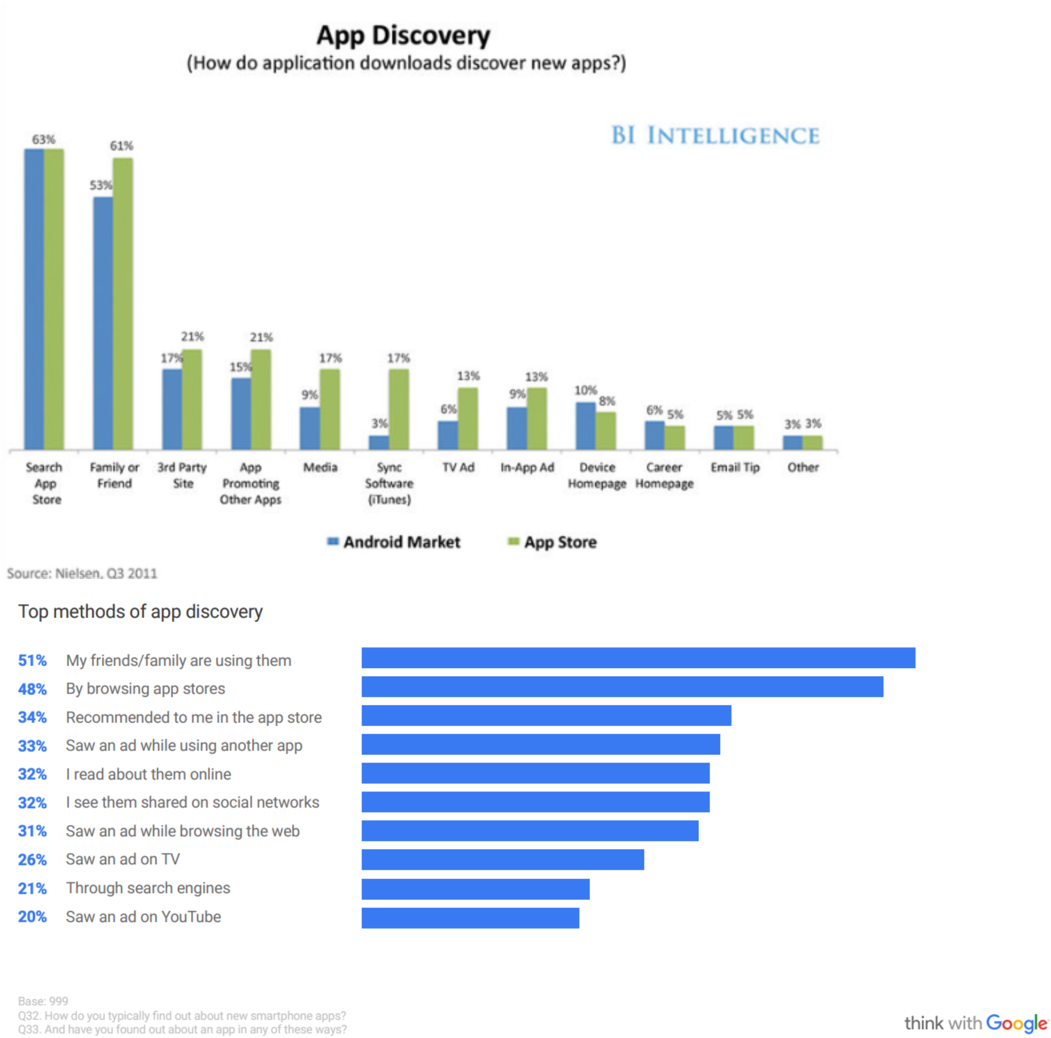


Figure A3 displays results from two additional app consumer surveys. The top panel shows results from a Nielsen survey from 2011. Over 60% of surveyed consumers on both Android and the iOS app store discover new products by searching the app store. This is not defined as "browsing" and could include using the search function of the store. It is a far more popular method of discovery than by going through other 3rd party sites or advertising.

The bottom panel of Figure A3 shows results from a 2016 Google and Nielsen survey. Once again, besides learning about apps from friends and family, the most popular discovery method is browsing the app store. This survey is from nearly two years after the end of my sample period and follows substantial improvements in the integration of apps into Google search results. Still, only 21% of users discover new apps using search engines.

Figure A3: Additional Google Play Consumer Product Discovery Surveys



B Appendix B

B.1 Data Management - Classifying App Types Using Text

I use a Random Forest machine learning algorithm that first maps the descriptions of the classified post-March 2014 apps into categories, and then projects this mapping on other apps. After removing "stopwords" (e.g., "and", "or") I convert app descriptions into vectors of words and then into term frequency-inverse document frequencies. This method assigns the highest weight to words appearing frequently in an app's description relative to the average description. I use the April 2014 apps as the training set for the Random Forest classifier (other classifiers such as KNN give similar result). I then apply the classifier to apps in every month prior to March 2014. This is similar to how Liu et al. (2014) map Google Play categories into Apple iTunes categories.

B.2 Data Management - Predicting App Downloads

Raw app data includes a range of cumulative downloads that an app accrues over its lifetime. The full list of download ranges is in Table B1. This range is observable in every snapshot of the store. It is conceptually straight-forward to define "per-period downloads" as the difference in lifetime downloads between period t and period $t-1$. For example, the difference in the lower bounds of lifetime downloads, or in the average of lifetime downloads.

Table B1: List of Cumulative Download Ranges

Lower Bound	Upper Bound
1	5
5	10
10	50
50	100
100	500
500	1,000
1,000	5,000
5,000	10,000
10,000	50,000
50,000	100,000
100,000	500,000
500,000	1 million
1 million	5 million
5 million	10 million
10 million	50 million
50 million	100 million
100 million	500 million
500 million	1 billion

The bandwidth increases with the number of downloads, starting at 4 downloads ([1-5], [5-10]) and increasing to 40 ([10-50]) and eventually to 400 million ([100 million - 500 million]). This introduces two possible sources of measurement error, which become worse for more successful apps: (1) overestimation of per period downloads for apps that move from one level to another. An app with a range of [100 thousand - 500 thousand] downloads that moves to the [500 thousand- 1 million] range in the next period could have been downloaded 500 thousand times or 3 times. (2) underestimation of per period downloads for successful apps. An app in the [100 million - 500 million] download range can have millions of downloads every week and stay in the same range.

I rely on two features of the data to recover weekly or monthly app downloads. First, the lifetime download bandwidth for *new entrants* is equal to the per-period bandwidth: an app that entered one period ago and is in the 10 thousand to 50 thousand range was downloaded between 10 thousand and 50 thousand times in the period. Second, I observe weekly category rankings which reflect the 500 most-downloaded apps in each category roughly over the past week.[1] At a weekly frequency, the rankings and downloads of new apps are known. Summary statistics are in Table B2.[2] I use these apps to predict the downloads of other apps in the market.

Several studies of online markets with best-seller lists find that the Pareto distribution accurately characterizes the rank-downloads relationship (Garg and Telang 2013; Chevalier and Mayzlin 2006).[3] The Pareto distribution is a negative exponential distribution where an app at rank n has exponentially more downloads than the app at rank $n+1$. I fit this distribution for every week and category by estimating an OLS regression of the logarithm of the rank of paid or non-paid ($p \in \{Paid, Non-Paid\}$) new app j in category c at week a of month t on the logarithm of the downloads for every category and week:

[1] It is not precisely known how the lists are determined, but Google releases (AdWeek.com) as well as anecdotal industry evidence (Quora) suggest that they reflect the downloads of apps over the previous several days.

[2] I can assign the lower bound of the bandwidth as the number of weekly downloads, the upper part of the bandwidth, or the average of the bandwidth. In the rest of the analysis of this paper I assign the lower end of the bandwidth, since the average and upper parts of the bandwidth produce unrealistic estimates of downloads.

[3] It is possible that the Pareto distribution does not correctly predict downloads in this market. Eeckhout (2004) shows that the Pareto distribution accurately predicts the rank-size relationship for the upper tail of the distribution but not necessarily for the lower tail. Liebowitz and Zentner (2020) similarly shows evidence of potential inaccuracy in approximations using distributional assumptions. I use an alternative measure of downloads that does not rely on the Pareto distribution in Online Appendix C.2.3. I also estimate my main results only for new apps, which are not affected by distributional assumptions. Results are qualitatively similar across the two sale proxies.

$$ln(\text{Downloads}_{jcat}) = \delta_{cp} + \delta_{tp} + \beta_1 ln(\text{Rank}_{jcat}) + \beta_2 ln(\text{Rank}_{jcat}) \times Paid_j + \mu_{jcat}$$

where δs are category and year/month dummies,[4] and where μ_{jcat} is a mean zero random variable representing measurement error. βs are slope coefficients that differ for paid and non-paid apps.[5] I use the lower bound of the bandwidth (minimum downloads in a week) as the dependent variable.[6] Summary statistics for new apps are in Table B2 and regression estimates are in Table B3. Estimated Pareto Distribution parameters are broadly consistent with similar exercises in the literature (Garg and Telang 2013; Leyden 2018).

I predict the downloads of all apps in the market with estimates from this regression. Only the top 500 ranks each week are observed. To generate rankings for the unranked apps, I sort them based on their number of cumulative lifetime downloads and their age in every week and break up ties by randomizing.[7]

This prediction algorithm depends on variation in app rankings over time. New apps should be able to enter into the rankings at different points in the distribution for me to estimate the Pareto relationship accurately. This is true in the data. While a large proportion of apps not change their rankings from week to week, many apps move at least two spots on a weekly basis. Figure B1 shows the distribution in weekly changes in game rankings.

Table B2: **Summary Statistics of New Apps at Weekly Level**

Variable	Mean	Std. Dev.	Min	Max	N Obs
Games					
Download Lower Bound	19,035	202,837	1	1 million	15,958
Non-Games					
Download Lower Bound	8,473	322,478	1	5 million	28,699

[4] δ_{cp} represents $\delta_{cp} \sum_c (D_c \times Paid_j + D_c \times (1 - Paid_j))$, where D_c is a dummy for whether j belongs to category c and $Paid_j$ is a dummy for whether app j is paid. δ_{tp} represents $\delta_{tp} \sum_t (D_t \times Paid_j + D_t \times (1 - Paid_j))$, where D_t is a dummy for month t.

[5] Predictions do not change substantially when slope coefficients also vary by time or category.

[6] Results using the upper bound or the average clearly overstate the number of downloads. For example, the model predicts each of the top 50 apps to have over 10 million weekly downloads.

[7] To check that randomization does not drive any of the main estimates, I re-estimate the analysis several times with different randomized seeds. The results remain qualitatively and quantitatively similar.

Table B3: **Regression Results on Downloads**

Outcome Variable:	*ln(Min Downloads Bound)*	
	(1)	(2)
	Games	Non-Games
ln(Rank)	-0.973***	-0.981***
	(0.017)	(0.011)
ln(Rank)×Paid	-1.170***	-1.016***
	(0.024)	(0.012)
Year/Month FE	•	•
Year/Month FE × Paid	•	•
Category FE	•	•
Category FE × Paid	•	•
Observations	15,958	28,699
R-squared	0.7529	0.8012

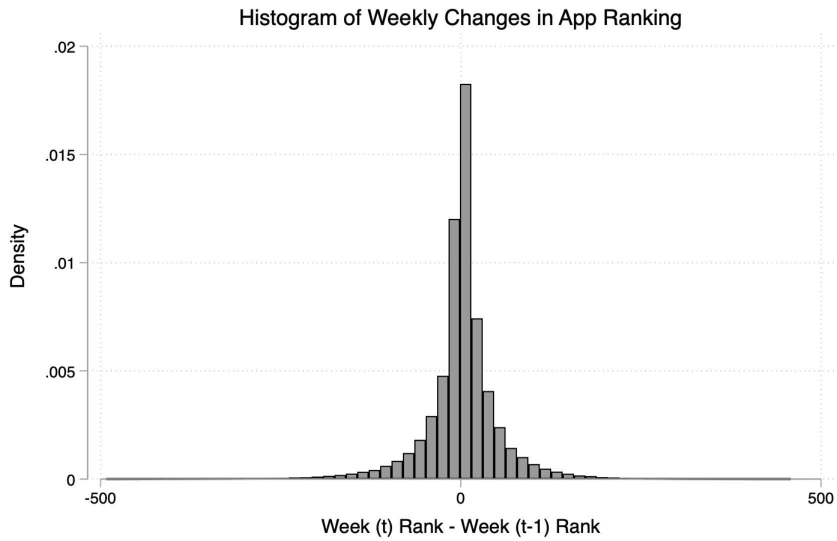
Notes: The sample in Column (1) are new games (games in their first week on the market). The sample in Column (2) are new non-games (non-game apps in their first week on the market). The outcome variable in both columns are the log of the lower bound of the number of weekly downloads for the apps. Controls include year/month and category fixed effects interacted with a paid app dummy. Heteroskedasticity robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

B.3 Summary Statistics

Table B4: **Summary Statistics at the App Type-Month Level**

Variable	**Mean**	**Std. Dev.**	**N**
Game Types			
Number of Apps	7,552	13,386	630
Number of New Apps	720	1,221	630
Non-Game Types			
Number of Apps	33,764	30,960	840
Number of New Apps	2,710	3,026	840

Figure B1: Weekly Changes in App Ranking on Top 500 Best-Seller Lists



C Appendix C

C.1 Downloads: March and April 2014 Estimates

In the main text, I use January and February 2014 as the "Pre"-policy period for the difference-in-differences download regressions in Tables 3 and 4. March and April are the "Post"-policy period. I do this because the policy change happened in the middle of March 2014, so data from that month may not fully reflect the policy change.

Table C1 replicates key regressions from Tables 3 and 4 from the main text using only data from February and March 2014. The "Pre" policy period is February 2014 and the "Post" policy period is March 2014. Results are qualitatively similar but quantitatively smaller than in the main text.

C.2 Downloads: Placebos, Paid Apps Only and Alternative Outcomes

C.2.1 Downloads: Placebo Time Periods

Most of the main results in Section 4.1 are computed using a restricted data sample of four months, comparing January and February 2014 to March and April 2014. A possible concern may be that estimated effects are not caused by re-categorization

Table C1: Downloads Difference in Differences: February and March Data

Outcome Variable:	(1) ln(Tot Type Dwnlds)	(2) ln(Downloads)	(3) ln(Tot Type Dwnlds)	(4) ln(Downloads)	(5) ln(Tot Type Dwnlds)	(6) ln(Downloads)	(7) Post/Pre Δ ln(Downloads)
Games × Post	1.081***	0.286**	0.251**	0.084*	0.144***	0.072***	
	(0.220)	(0.132)	(0.113)	(0.050)	(0.012)	(0.009)	
Games × Post × No Pre-Existing			1.494***	1.791***			
			(0.242)	(0.082)			
Games × Post × Small Type					0.632***	0.921***	
					(0.010)	(0.119)	
Post/Pre Δ ln(N Apps)							-0.594***
							(0.032)
Unit of Observation	App-Type	App	App-Type	App	App-Type	App	App
Sample Period	Feb 14/Mar 14	Feb 14/Mar 14	Feb 14/Mar 14	Feb 14/Mar 14	Feb 14/Mar 14	Feb 14/Mar 14	Feb 14/Mar 14
Sample	All	All	All	All	All Non-Games + Card, Casino, Arcade and Action	All Non-Games + Card, Casino, Arcade and Action	All Games
Observations	84	2,574,302	84	2,574,302	56	2,330,302	142,419
R-squared	0.833	0.981	0.944	0.989	0.998	0.992	0.868
Year/Month FE	•	•	•	•	•	•	
App-Type FE	•		•		•		
App FE		•		•		•	

Notes: The sample period in all columns is February and March 2014. Data in Columns (1), (3) and (5) consists of monthly observations at the app-type level. Data in Columns (2), (4), (6) and (7) consists of monthly observations at the app level. Columns (1)-(4) include all apps. Columns (5) and (6) include all non-game apps and Arcade, Action, Card and Casino game apps. Column (7) includes all game apps. Outcomes for Columns (1)-(6) are the natural logarithms of downloads at each aggregation level. The outcome for Column (7) is the difference between the natural log of app downloads in March 2014 and downloads in February 2014. Controls include year and month fixed effects and app-type or app fixed effects. Column (7) does not have fixed effects because it is a cross sectional regression in first-differences. Additional app-level controls include average app ratings, a dummy for whether an app is free or paid, the price of paid apps and app-age specific fixed effects. The variable "Games × Post" is a dummy variable equal to 1 for games, or game app-types in March 2014. "No Pre-Existing" is a dummy variable equal to 1 for apps or app-types with no pre-existing categories before March 2014. "Small Split" is a dummy variable equal to 1 for Action and Casino games. "Post/Pre ln(N Apps in Category)" is the difference in the natural log of the number of apps in the category of app j in March 2014 and the number of apps in February 2014. Standard errors are clustered at the app-type level for Columns (1)-(6) and are robust to heteroskedasticity in Column (7). *** p<0.01, ** p<0.05, * p<0.1

but by diverging time-trends between games and non-games or across game types. To test whether this is the case, I re-estimate the regressions using two comparable time periods without a re-categorization event: November 2013 to February 2014, and March 2014 to June 2014. For each sample, I estimate the effects of a non-existent re-categorization event: between December and January for the first sample, and between April and May for the second sample.

The first "placebo" sample verifies that download trends between games and non-games or across game types were not diverging before re-categorization took place. It also helps test whether it was the actual re-categorization or the *announcement* of re-categorization in December 2013 changed download outcomes. If changes in downloads were actually caused by re-categorization and improved consumer discovery, there is no reason to expect statistically significant differences in downloads before. The second "placebo" sample further verifies the effects of re-categorization. The policy was a permanent event - a consumer in May 2014 should have had as easy a time finding the "Music" game category as a consumer in April 2014. If re-categorization improved consumer discovery technology, these improvements should be locally persistent over time.[8]

I show results using alternative time periods in Table C2. These estimates replicate the main specifications shown in Tables 3 and 4.[9] Panel (a) shows results using the Nov 2013 - Feb 2014 sample and panel (b) shows results using the March 2014 - June 2014 sample.

Nearly all estimates for the alternative time periods are statistically null. They are also generally substantially smaller in magnitude than estimates in the main text. There is some evidence of heterogeneous time-trends across game types without pre-existing categories and game types with pre-existing categories prior to re-categorization in Columns (3) and (4) in panel (a). However, relative to the baseline group of non-game types or apps, the total change in downloads for game app-types without pre-existing categories is still null. For both Columns (3) and (4), the sum of the "Games × Placebo Post" and "Games × Placebo Post × No Pre-Existing" is not statistically significantly different than zero. These results show that differential changes in downloads between games and non-games occurred only during re-categorization. These results also show that the effects of re-categorization on downloads are persistent. Downloads four months after re-categorization were not statistically significantly different than the month after re-categorization. This suggests that the re-categorization event directly caused the change in downloads, consistent with permanent reductions in search costs for consumers due to an improvement in category informativeness and a reduction in the number of apps per category.

C.2.2 Downloads: Paid Apps

Downloads of free apps may not be accurate proxies of consumer app usage as consumers can easily install and uninstall such apps from their phones without fully inspecting or using them. This is not the case for paid apps that consumers are required to spend money on upfront. I re-estimate the main regressions from Tables 3 and 4 in the main text after restricting the sample to paid apps. I also test for placebo effects using alternative time periods as in Table C2.

Estimates for the sample of paid apps are in Table C3. Odd columns show

[8] Over a longer period of time, changes in product assortment and entry may introduce additional congestion costs into the market, mitigating some of the immediate decreases in discovery costs. The changes in category informativeness, however, should be very persistent over time.

[9] Full time-varying estimates of treatment effects for specifications where I use data from January 2012 to December 2014 are in Figure C1. They also show the main download effects appear only following the actual re-categorization.

Table C2: Downloads Difference in Differences: Alternative Time Periods

Outcome Variable:	(1) ln(Type Downloads)	(2) ln(Downloads)	(3) ln(Type Downloads)	(4) ln(Downloads)	(5) ln(Type Downloads)	(6) ln(Downloads)
			Panel (a): Nov 2013 - Feb 2014			
Games × Placebo Post	-0.131	-0.291	-0.219	-0.302	-0.210	-0.312
	(0.137)	(0.199)	(0.154)	(0.198)	(0.132)	(0.149)
Games × Placebo Post × No Pre-Existing			0.158**	0.121**		
			(0.027)	(0.035)		
Games × Placebo Post × Small Type					0.019	0.105
					(0.039)	(0.111)
Observations	168	4,892,297	168	4,892,297	112	4,427,584
R-squared	0.987	0.978	0.987	0.978	0.968	0.982
			Panel (b): March 2014 - June 2014			
Games × Placebo Post	0.223	0.065	0.110	0.056	0.124	0.079
	(0.202)	(0.140)	(0.120)	(0.117)	(0.107)	(0.106)
Games × Placebo Post × No Pre-Existing			0.204	0.064		
			(0.147)	(0.168)		
Games × Placebo Post × Small Type					0.082	-0.030
					(0.074)	(0.119)
Observations	168	5,496,539	168	5,496,539	112	4,935,476
R-squared	0.900	0.985	0.909	0.985	0.955	0.986
Unit of Observation:	App-Type	App	App-Type	App	App-Type	App
Sample	All	All	All	All	All Non-Games + Card, Casino, Arcade and Action	All Non-Games + Card, Casino, Arcade and Action
Year/Month FE	•	•	•	•	•	•
App-Type FE	•		•		•	
App FE		•		•		•

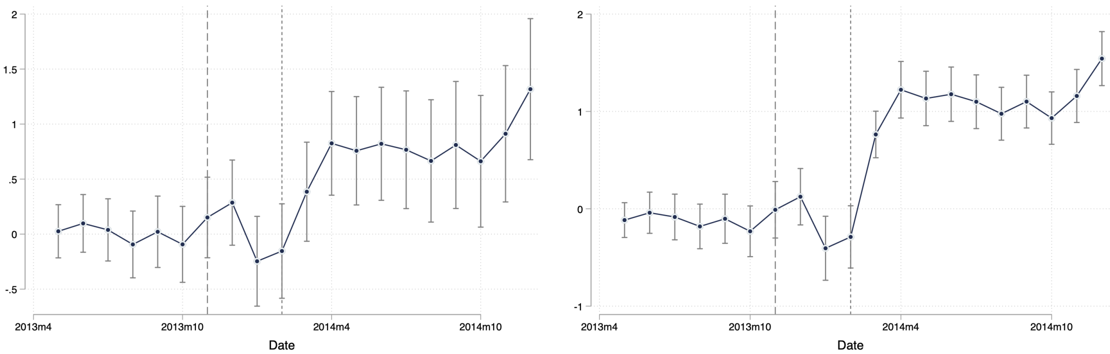
Notes: The sample period in panel (a) covers November 2013 to February 2014. The sample period in panel (b) covers March 2014 to June 2014. Sample in Columns (1), (3) and (5) consists of monthly observations at the app-type level. Sample in Columns (2), (4) and (6) consists of monthly observations at the app level. Outcomes are natural logarithms of downloads at each aggregation level. Controls include year and month fixed effects, app-type fixed effects, or app fixed effects, depending on the column. Additional app-level controls for Columns (2), (4) and (6) include average app ratings, a dummy for whether the app is free or paid, the price of the app if it is paid, and app age-specific fixed effects. The variable "Games × Placebo Post" is a dummy variable equal to 1 for games, or game app-types during and after January 2014 for panel (a) and during and after May 2014 for panel (b). Standard errors are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

estimates from aggregated app-type regressions with app-type fixed effects. Even columns show estimates from app-level regressions and even columns are at the app level with app fixed effects. Panel (a) shows estimates of the regressions for the January 2014 - April 2014 period, testing for the effects of a re-categorization event in March. Panel (b) shows estimates using the November 2013 - February 2014 period, testing the effects of a non-existent re-categorization event between December and January. Panel (c) similarly shows estimates using the March 2014 - June 2014 period with a non-existent re-categorization event between April and May.

In each panel the first two columns show the baseline average effects. The next two columns test for heterogeneity across game types that had pre-existing categories before the policy and those that did not. Such heterogeneity identifies changes in discovery costs through increasing informativeness. The last two columns test for heterogeneity across Arcade, Action, Card and Casino game types, where Action and

Figure C1: Downloads Timing Tests

(a) Outcome: ln(Total Downloads) (b) Outcome: ln(Total Type Downloads)



Notes: Each panel shows estimates of coefficients τ_t from Equation 2 at different aggregation levels. Panel (a) is estimated at the game/non-game level and panels (b) is estimated at the app-type level. Data from January 2012 to December 2014 is used throughout. Additional controls in each regression include year/month fixed effects, game/non-game or app-type fixed effects, and game/non-game or app-type specific trends. Standard errors for panel (a) are robust to heteroskedasticity and standard errors for panel (b) are clustered at the app-type level. 95% confidence intervals shown. In each panel, the first dashed vertical line represents the announcement of re-categorization and the second dashed vertical line represents the start of the re-categorization period.

Casino were much smaller before re-categorization. Such heterogeneity should identifies changes in discovery costs through reducing the number of apps per category and congestion.

Results are consistent with those in the main text and the robustness checks above. Panel (a) shows that on average, downloads for paid games increased over non-games after re-categorization. Paid games belonging to types without pre-existing categories and games belonging to types with fewer apps are driving the main effects. Estimates are statistically significant at the 95% confidence level and are larger than in the main text. Results for the two "placebo" events, before and after the actual re-categorization show statistically null effects. These estimates confirm that consumer discovery costs fell in response to re-categorization.

C.2.3 Downloads: Alternative Outcomes

I use three alternative outcome variables to test the robustness of estimates in Section 4.1. Two of the alternative outcomes do not rely on the procedure described in Appendix B.2. As discussed by Liebowitz and Zentner (2020), the parametric assumptions used to generate most monthly download values in Appendix B.2 can

Table C3: Downloads Difference-in-Differences: Paid App Sample

Outcome Variable:	(1) ln(Type Dwnlds)	(2) ln(Dwnlds)	(3) ln(Type Dwnlds)	(4) ln(Dwnlds)	(5) ln(Type Dwnlds)	(6) ln(Dwnlds)
			Panel (a): Jan 2014 - Apr 2014			
Games × Post	1.536***	0.489**	0.594**	0.252***	0.332**	0.255***
	(0.228)	(0.102)	(0.163)	(0.040)	(0.069)	(0.020)
Games × Post × No Pre-Existing			1.697**	2.178***		
			(0.315)	(0.220)		
Games × Post × Small Type					1.322**	1.560***
					(0.237)	(0.154)
Observations	168	972,440	168	972,440	112	883,472
R-squared	0.761	0.978	0.876	0.986	0.953	0.990
			Panel (b): Nov 2013 - Feb 2014			
Games × Placebo Post	0.052	-0.024	0.064	-0.029	-0.038	-0.022
	(0.034)	(0.057)	(0.043)	(0.057)	(0.056)	(0.045)
Games × Placebo Post × No Pre-Existing			-0.022	0.049*		
			(0.065)	(0.019)		
Games × Placebo Post × Small Type					0.314*	0.020
					(0.126)	(0.043)
Observations	168	942,428	168	942,428	112	858,336
R-squared	0.990	0.989	0.990	0.989	0.989	0.989
			Panel (c): Mar 2014 - Jun 2014			
Games × Placebo Post	0.255	0.103	0.142	0.090	0.142	0.093
	(0.138)	(0.069)	(0.084)	(0.072)	(0.084)	(0.052)
Games × Placebo Post × No Pre-Existing			0.203	0.096		
			(0.162)	(0.196)		
Games × Placebo Post × Small Type					0.104	0.023
					(0.045)	(0.101)
Observations	168	958,186	168	958,186	112	866,545
R-squared	0.925	0.990	0.934	0.990	0.963	0.990
Unit of Observation:	App-Type	App	App-Type	App	App-Type	App
Sample	All Paid	All Paid	All Paid	All Paid	All Paid Non-Games + Paid Card, Casino, Arcade and Action	All Paid Non-Games + Paid Card, Casino, Arcade and Action
Year/Month FE	•	•	•	•	•	•
App-Type FE	•		•		•	
App FE		•		•		•

Notes: The sample throughout all panels and columns only includes *paid* apps with non-zero prices. The sample period in panel (a) covers January 2014 to April 2014. The sample period in panel (b) covers November 2013 to February 2014. The sample period in panel (c) covers March 2014 to June 2014. Sample in Columns (1), (3) and (5) consists of monthly observations at the app-type level. Sample in Columns (2), (4) and (6) consists of monthly observations at the app level. Outcomes are natural logarithms of downloads at each aggregation level. Controls include year and month fixed effects, app-type fixed effects, or app fixed effects, depending on the column. Additional app-level controls for Columns (2), (4) and (6) include average app ratings, the price of the app, and app age-specific fixed effects.The variable "Games × Post" is a dummy variable equal to 1 for games, or game app-types during and after March 2014 for panel (a). The variable "Games × Placebo Post" is a dummy variable equal to 1 for games, or game app-types after during and after January 2014 for panel (b) and during and after May 2014 for panel (c). Standard errors are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

produce biased estimates of actual downloads.

The first alternative outcome restricts the sample of apps to new apps: apps that entered the store in month t. For these apps, the approximation bias is minimal since they are used to fit the model in Appendix B.2.

The second alternative outcome is a simpler proxy for monthly downloads: the difference in the number of user ratings for an app between two periods. For app

j, downloads for period t are approximated by the number of user ratings in period t minus the number of user ratings in period $t-1$ (N Ratings_{jt} − N Ratings_{jt-1}).[10] This proxy relies on a simple intuitive relationship - if a certain proportion of users who download an app also rate it, apps with more downloads will also have more ratings. Ratings on Google Play have to come from downloads, and it is unlikely that users will wait over a month to rate an app they downloaded. Such proxies have been used previously in papers studying mobile apps, such as Kummer and Schulte (2019). This approach has limitations as the relationship between downloading and rating is not necessarily strictly monotonic in the number of downloads. Some types of apps may be very frequently downloaded but not frequently rated, whereas other apps are both frequently downloaded and rated. The proportion of users who rate apps may also decrease in app popularity. This would create a bias in the download proxy. For this reason I choose to use downloads calculated according to B.2 as the main specification.

The last alternative outcome is the absolute number of predicted downloads rather than the natural log of downloads (using the predicted measure of downloads from Online Appendix B.2).

Results using these three outcomes are in Table C4. There are four panels in the table. For ease of comparison, Panel (a) provides results using the baseline outcome from the main text. Panel (b) shows results using only new apps. Panel (c) shows results using the difference in the number of ratings to approximate downloads. Panel (d) shows results using the absolute number of predicted downloads. Odd columns aggregate data at the app-type level and even columns use app level data. App level fixed effects are included for regressions in panels (a), (c) and (d). Panel (b) only includes app-type fixed effects as I only observe each new app once. App level regressions in each panel are done using all non-game apps and game apps belonging to app-types without categories before the policies. I pick this sample as discovery costs should fall for this set of game apps (see panel (a) of Table 4). Columns (1) and (2) use the baseline January 2014 - April 2014 four month sample period. Columns (3) and (4) use November 2013 to February 2014 as the sample period, with a “placebo” event between December 2013 and January 2014. Columns (5) and (6) use March 2014 to June 2014 as the sample period, with a “placebo” event between April 2014 and May 2014.

Results are qualitatively equivalent to the main specification. Results for ratings based downloads are different in magnitude than in the main text because of the

[10] On occasion, ratings and reviews disappear from the Google Play store for various reasons including service term violations and the number of ratings falls between period $t-1$ and period t. This occurs for less than 0.7% of observations. I bound changes in ratings to zero from below.

different definition. However, downloads statistically significantly increase for games relative to non-games in March and April 2014 relative to January and February. The same does not happen in May and June 2014 relative to March and April, or in January and February 2014 relative to November and December 2013. Estimates in panel (b) using the sample of new apps are also similar to those in the main text.

Table C4: Downloads Difference in Differences Estimates: Alternative Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel (a): Baseline			
Outcome Variable:	ln(Tot. Type Downloads)	ln(App Downloads)	ln(Tot. Type Downloads)	ln(App Downloads)	ln(Tot. Type Downloads)	ln(App Downloads)
Games × Post	1.399**	2.214***				
	(0.288)	(0.302)				
Games × Placebo Post			-0.131	-0.174		
			(0.137)	(0.173)		
Games × Placebo Post					0.223	0.120
					(0.202)	(0.273)
Observations	168	4,646,394	168	4,294,910	168	4,825,839
R-squared	0.780	0.982	0.987	0.982	0.900	0.986
			Panel (b): New App Sample			
Outcome Variable:	ln(Tot. New Type Downloads)	ln(New App Downloads)	ln(Tot. New Type Downloads)	ln(New App Downloads)	ln(Tot. New Type Downloads)	ln(New App Downloads)
Games × Post	1.840**	2.440***				
	(0.318)	(0.128)				
Games × Placebo Post			0.486	-0.315		
			(0.207)	(0.172)		
Games × Placebo Post					-0.045	-0.087
					(0.085)	(0.120)
Observations	168	378,987	168	481,961	168	363,619
R-squared	0.821	0.622	0.849	0.507	0.931	0.706
			Panel (c): Ratings Based Downloads			
Outcome Variable:	ln(Tot. Type Δ Ratings)	ln(App Δ Ratings)	ln(Tot. Type Δ Ratings)	ln(App Δ Ratings)	ln(Tot. Type Δ Ratings)	ln(App Δ Ratings)
Games × Post	0.092**	0.196*				
	(0.025)	(0.062)				
Games × Placebo Post			0.068	0.172		
			(0.032)	(0.147)		
Games × Placebo Post					0.072	0.050
					(0.039)	(0.062)
Observations	168	4,646,680	168	4,284,464	168	4,829,472
R-squared	0.992	0.918	0.987	0.877	0.991	0.917
			Panel (d): Absolute Downloads			
Outcome Variable:	Tot. Type Downloads	App Downloads	Tot. Type Downloads	App Downloads	Tot. Type Downloads	App Downloads
Games × Post	1301604.410*	739.440**				
	(459,123.970)	(227.632)				
Games × Placebo Post			-217,949.749	-198.974		
			(313,913.912)	(106.365)		
Games × Placebo Post					417,688.494	110.491
					(383,730.890)	(184.156)
Observations	168	4,646,394	168	4,294,910	168	4,825,839
R-squared	0.854	0.780	0.843	0.817	0.908	0.931
Unit of Observation:	App-Type	App	App-Type	App	App-Type	App
Sample Period:	Jan 14/Apr 14	Jan 14/Apr 14	Nov 13/Feb 14	Nov 13/Feb 14	Mar 14/Jun 14	Mar 14/Jun 14
Sample	All	All Non-Games + Games w/o Pre-Exist Cats.	All	All Non-Games + Games w/o Pre-Exist Cats.	All	All Non-Games + Games w/o Pre-Exist Cats.
Year/Month FE	•	•	•	•	•	•
App-Type FE	•		•		•	
App FE		•		•		•

Notes: Sample for odd columns includes all apps and for even columns includes all non-games and Adventure, Board, Education, Family, Music, Role Playing, Simulation, Strategy, Trivia and Word games. Sample period in Cols (1)-(2) covers Jan/Apr 2014. Sample period in Cols (3)-(4) covers Nov 2013 - Feb 2014. Sample period in Cols (5)-(6) covers Mar/June 2014. Odd columns sample Outcomes are defined as the title of each column/panel combination. Controls include year and month fixed effects, app-type fixed effects, or app fixed effects, depending on the column. Additional app-level controls for even columns include average app ratings, a dummy for whether the app is free or paid, the price of the app if it is paid, and app age-specific fixed effects. Panel (b) does not include app-level fixed effects. In Cols (1)-(2) "Games × Post" is a dummy variable equal to 1 for games, or game app-types during and after March 2014. In Cols (3)-(4) "Games × Placebo Post" is a dummy variable equal to 1 for games, or game app-types during and after January 2014. In Cols (5)-(6) "Games × Placebo Post" is a dummy variable equal to 1 for games, or game app-types during and after May 2014. Standard errors are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

C.3 Downloads: Changes in Number of Apps per Category

This section shows how, for a given app, the number of other apps in its category changes between November/December 2013 and January/February 2014, January/February 2014 and March/April 2014, and March/April 2014 and May/June 2014. I do this by first calculating, for each app, the average number of other apps in its category in each two month period.[11] Then I calculate, for each app, the difference between two successive periods.

The distribution of changes appears in Figure C2. This figure has three panels representing the three sets of changes I examine. The distribution of changes in panels (a) and (c) is entirely different than the distribution of changes in panel (b). In panels (a) and (c), the number of other apps in the category of an app increase. This is consistent with the general growth in the number of apps on Google Play over time. Panel (b) shows that between Jan/Feb and Mar/Apr 2014, all apps experienced a drop in the number of other apps in their category. The variation in this drop represents differences between what category the app belonged to before re-categorization and its category/app-type after re-categorization.

C.4 Downloads and Long Run Entry

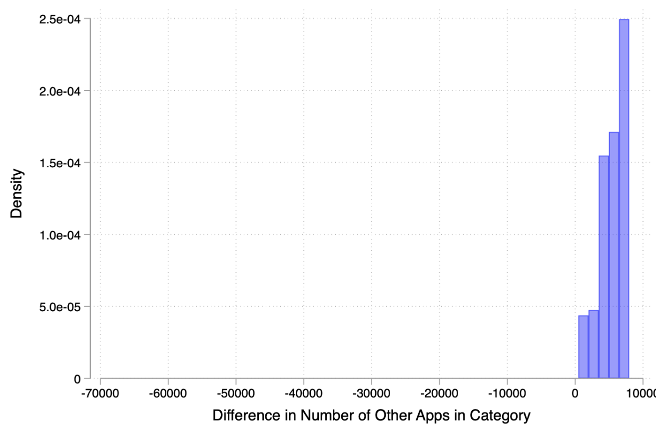
In Columns (7) and (8) in Table 4 in the main text I show evidence of the effects of short-run changes in the number of apps in an app's category on app-level downloads. I do this using short-run changes induced by re-categorization, which move apps from broad categories to narrow categories that reflect their app types. Since where the apps end up is primarily determined by their pre-existing app-type and entry does not change much, this primarily reflects congestion rather than changes in the competition intensity (as reflected by the number of substitutes) for each app.

However, this does not necessarily means that longer run changes in the number of apps of each app-type will have similar effects. As mentioned above, increases in the number of apps of each app-type could affect both competition intensity and congestion. I test for this in the data by relating long run differences in app-type entry to long run differences in downloads in the post re-categorization period. For a given app j, I calculate the difference in their downloads between December 2014 and March 2014. I then regress this difference on the difference in the number of apps of their type (which also coincides with their category) between December 2014

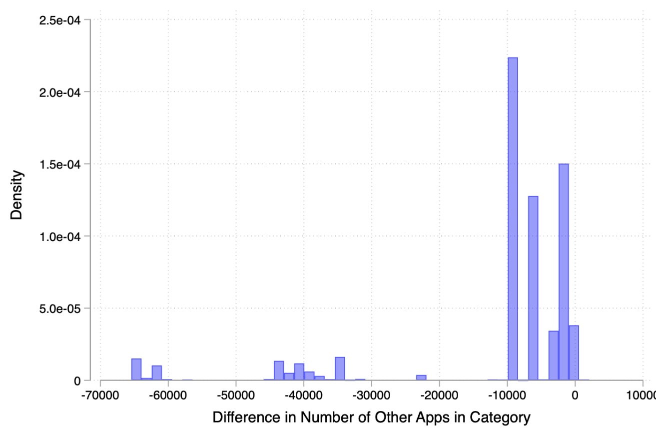
[11] I refer to categories here rather than app types since consumers use the stated category structure to search. See Section 2.1 for additional discussion. I use two month periods since these are comparable to the time periods I use in Table 4.

Figure C2

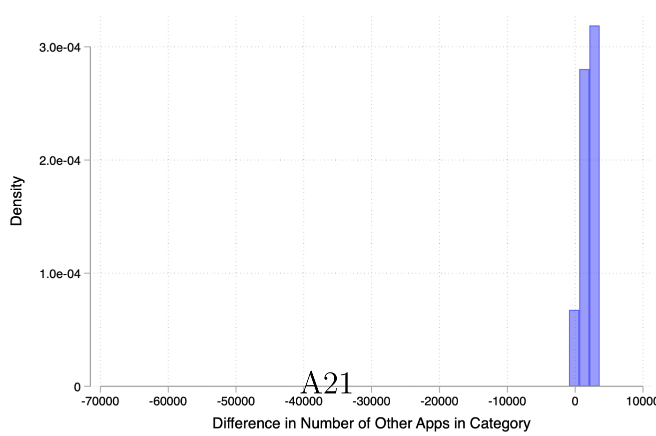
(a) ln(N Apps Jan/Feb 2014) - ln(N Apps Nov/Dec 2013)



(b) ln(N Apps Mar/Apr 2014) - ln(N Apps Jan/Feb 2014)



(c) ln(N Apps May/Jun 2014) - ln(N Apps Mar/Apr 2014)



Notes: Each panel shows the distribution of changes in app-level changes in the number of other apps in their category over time. For each app j, I calculate the difference in the natural log of the number of apps in their category between two successive periods. If app j is in category c in period t and category c' in period $t+1$, the difference is $ln(N_{c't+1}) - ln(N_{ct})$ In panel (a), the difference is between Jan/Feb 2014 (on average) Nov/Dec 2013. In panel (b), the difference is between Jan/Feb 2014 and Mar/Apr 2014. In panel (c), the difference is between Mar/Apr 2014 and May/Jun 2014.

and March 2014. The estimating equation is as follows:

$$ln(Downloads_{j,Dec}) - ln(Downloads_{j,Mar}) = \alpha(ln(NApps_{jc,Dec}) - ln(NApps)_{j,Mar}) + \beta X_j + \epsilon_j \quad (1)$$

where X_j are app characteristics I control for to account for unobservable heterogeneity not fully absorbed by the within-app differencing.

Estimates of this regression are in Table C5. Column (1) shows the baseline results and Column (2) interacts the difference in the number of apps with a dummy for app-types that did not have pre-existing categories before the policy. The coefficient on the difference in the number of apps in the category is -0.66, suggesting that a one percent increase in the number of apps reduces app downloads by 0.66%. Column (2) shows that there is minimal heterogeneity in this coefficient across different app-types.

Table C5: Long Run Changes in Downloads and Entry

	(1)	(2)
Outcome:	Dec/Mar Δ ln(Downloads)	Dec/Mar Δ ln(Downloads)
Dec/Mar Δ ln(N Apps)	-0.655***	-0.629***
	(0.016)	(0.021)
Dec/Mar Δ ln(N Apps) × No Pre-Existing		-0.036*
		(0.020)
Unit of Observation	App	App
Sample	All Games	All Games
Sample Period	Mar and Dec 2014	Mar and Dec 2014
App Controls	•	•
Observations	121,134	121,134
R-squared	0.556	0.556

Notes: Sample includes all apps present in both March and December 2014. App-level controls include average app ratings, a dummy for whether the app is free or paid, the price of the app if it is paid, and app age-specific fixed effects. Standard errors are robust to heteroskedasticity. *** p<0.01, ** p<0.05, * p<0.1

These long-run estimates of the elasticity between changes in the number of apps in a category and app-level downloads are strikingly similar to short-run estimates in the main text. The coefficient on the short run re-categorization driven changes in the number of apps on the number of downloads is -0.65. This is reassuring, since the short-run re-categorization effect on congestion I identify in the main text seems to also be operating in the longer run. If changes in longer run entry were also generating pressure on app downloads through competition from additional substitutes, I would expect the coefficient in the long run regression to be substantially larger (in absolute terms). One possibility is that competition is already intense in the app market in March 2014 such that additional entry between March and December does not significantly increase it.

The OLS regressions estimated in this section may be subject to endogeneity concerns as both downloads and app-type entry could be determined by common unobservable shocks. However, in many ways the evolution of app-type entry after re-categorization is driven by changes caused by re-categorization. In Section 4.2 I show that changes in entry after re-categorization are driven by whether the app-type's discovery costs fell during re-categorization. This means that app entry in November 2014 (e.g., between November and December) was largely driven by the re-categorization which happened eight months prior.

C.5 Downloads: Heterogeneity by App Entry Cohort and by App Popularity

In this section, I decompose the average effects of re-categorization on game app downloads and test which apps benefit the most from re-categorization. I show evidence along two dimensions: app entry cohort and past app popularity. Both of these matter for understanding the benefits of re-categorization.

The first set of regressions split the overall sample to compare how downloads change for a game older than six months at the time of re-categorization as compared to a game that is younger than six months (all relative to the baseline of similar non-game apps which do not experience re-categorization). Estimates are in Table C6.

Table C6: Downloads Difference in Differences Heterogeneity by App Age

	(1) App Age $\geq$ 6 months	(2) App Age $<$ 6 months
Outcome:	N Monthly Downloads	N Monthly Downloads
Games × Post	1,180**	481**
	(348)	(151)
Unit of Observation	App	App
Sample Period	Jan 14/Apr 14	Jan 14/Apr 14
Sample	All Non-Games + Games w/o Pre-Exist Cats.	All Non-Games + Games w/o Pre-Exist Cats.
App FE	•	•
Year/Month FE	•	•
App Controls	•	•
Observations	2,615,204	2,031,190
R-squared	0.813	0.679

Notes: Sample period for all columns is January 2014 - April 2014. Sample in Column (1) includes all non-games and games without pre-existing categories that have been in the store for six months or longer in period t. Sample in Column (2) includes all non-games and games without pre-existing categories that have been in the store for less than six months in period t. App-level controls include average app ratings, a dummy for whether the app is free or paid, the price of the app if it is paid, and app age-specific fixed effects. "Games × Post" is a dummy variable equal to 1 for games, or game app-types during and after March 2014. Standard errors are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

The second set of regressions compare how re-categorization affects apps with fewer than 1,000 past downloads as compared to apps with more than 1,000 but fewer than 10,000 past downloads, as compared to apps with more than 10,000 past downloads. Results are in Table C7.

Table C7: Downloads Difference in Differences Heterogeneity by Past App Popularity

	(1) Apps w/ <1,000 Cumulative Downloads	(2) Apps w/ ≥1,000 & <10,000 Cumulative Downloads	(3) Apps w/ ≥10,000 Cumulative Downloads
Outcome:	N Monthly Downloads	N Monthly Downloads	N Monthly Downloads
Games × Post	226**	502**	1,529*
	(60)	(142)	(495)
Unit of Observation	App	App	App
Sample Period	Jan 14/Apr 14	Jan 14/Apr 14	Jan 14/Apr 14
Sample	All Non-Games + Games w/o Pre-Exist Cats.	All Non-Games + Games w/o Pre-Exist Cats.	All Non-Games + Games w/o Pre-Exist Cats.
App FE	•	•	•
Year/Month FE	•	•	•
App Controls	•	•	•
Observations	3,123,885	830,596	599,227
R-squared	0.739	0.773	0.786

Notes: Sample period for all columns is January 2014 - April 2014. Sample in Column (1) includes all non-games and games without pre-existing categories that have been fewer than 1,000 cumulative downloads in period t. Sample in Column (2) includes all non-games and games without pre-existing categories with more than 1,000 cumulative downloads but fewer than 10,000 downloads in period t. Sample in Column (3) includes all non-games and games without pre-existing categories with more than 10,000 cumulative downloads in period t. App-level controls include average app ratings, a dummy for whether the app is free or paid, the price of the app if it is paid, and app age-specific fixed effects. "Games × Post" is a dummy variable equal to 1 for games, or game app-types during and after March 2014. Standard errors are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

Estimates from Tables C6 and C7 show that older apps benefit more from re-categorization and the reduction of discovery costs than younger apps. However, apps with more downloads progressively benefit more from the re-categorization as well. This is the result of the underlying structure of the store: category spaces with best-seller and featured app lists. The re-categorization created additional spaces which benefit some apps that would have previously been excluded. But since more popular apps have a higher chance of appearing on various lists, the apps that benefit are more popular ones. Apps that are very popular and were already featured in category lists before re-categorization also benefit since there are informational benefits and consumers are better able to reach them. Overall, the distribution of effects is consistent with the structure of the app store.

C.6 Entry and Quality: App-Type Specific Difference in Differences Estimates

Table C8: **Entry and Quality: App-Type Difference-in-Differences Estimates**

Outcome Variable:	ln(N Type Entrants)	Mean Type Entrant Rating
Action × Post	1.079***	-0.026**
	(0.102)	(0.012)
Adventure × Post	1.207***	-0.093***
	(0.071)	(0.019)
Arcade × Post	-0.047	0.210***
	(0.077)	(0.015)
Board × Post	0.918***	-0.009
	(0.057)	(0.015)
Card × Post	-0.125*	-0.054***
	(0.065)	(0.013)
Casino × Post	0.415***	-0.173***
	(0.085)	(0.013)
Casual × Post	0.051	-0.050*
	(0.052)	(0.026)
Educational × Post	1.040***	-0.048***
	(0.069)	(0.013)
Family × Post	0.649***	-0.096***
	(0.058)	(0.015)
Music × Post	0.658***	-0.100***
	(0.056)	(0.028)
Puzzle × Post	-0.208***	0.003
	(0.072)	(0.020)
Racing × Post	-0.039	0.049***
	(0.054)	(0.013)
Role Playing × Post	0.476***	0.004
	(0.057)	(0.014)
Simulation × Post	0.910***	-0.235***
	(0.054)	(0.018)
Sports × Post	0.882***	-0.079***
	(0.067)	(0.025)
Strategy × Post	0.741***	-0.124***
	(0.078)	(0.015)
Trivia × Post	0.730***	-0.043***
	(0.064)	(0.012)
Word × Post	0.678***	-0.226***
	(0.062)	(0.018)
Observations	1,470	1,470
R-squared	0.977	0.686

Notes: Sample period for all columns is January 2012 - December 2014. All game and non-game app-types included. Each observation is an app-type month. The outcome variable in Column (1) is the natural logarithm of the number of new apps in an *app-type*. The outcome variable in Column (2) is the average rating of new apps in an *app-type*. Controls include app-type and date fixed effects and app-type specific time trends. Standard errors are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

Table C9: Entry and Quality Difference in Differences Estimates with Alternative Outcomes

	Panel (a): Absolute Number of Entrants			
Outcome:	N Entrants	N Type Entrants	N Type Entrants	N Type Entrants
Games × Post	34,658.448**	1,463.367***	1,134.432***	854.808**
	(16,259.260)	(310.944)	(394.645)	(348.410)
Games × Post × No Pre-Existing			592.083**	
			(296.673)	
Games × Post × Small Type				1,214.406***
				(438.110)
Observations	70	1,470	1,470	980
R-squared	0.876	0.775	0.776	0.776
	Panel (b): Mean Entrant Share of 1 Star Ratings			
Outcome:	Mean Share	Mean Type Share	Mean Type Share	Mean Type Share
Games × Post	0.005	0.017***	0.005	-0.013
	(0.005)	(0.005)	(0.007)	(0.013)
Games × Post × No Pre-Existing			0.021**	
			(0.008)	
Games × Post × Small Type				0.040**
				(0.016)
Observations	70	1,470	1,470	980
R-squared	0.970	0.639	0.642	0.747
	Panel (c): Mean Entrant ln(MB Size)			
Outcome:	Mean ln(MB)	Mean Type ln(MB)	Mean Type ln(MB)	Mean Type ln(MB)
Games × Post	-0.600	-7.758***	-8.100**	-2.404
	(0.928)	(1.977)	(3.996)	(1.453)
Games × Post × No Pre-Existing			0.615	
			(4.264)	
Games × Post × Small Type				-24.737**
				(9.947)
Observations	70	1,470	1,470	980
R-squared	0.964	0.609	0.609	0.680
	Panel (d): Download Weighted Entrant Avg. Rating			
Outcome:	Mean W. Rating	Mean W. Type Rating	Mean W. Type Rating	Mean W. Type Rating
Games × Post	14.129	-0.522***	-0.290*	0.305*
	(47.128)	(0.119)	(0.150)	(0.160)
Games × Post × No Pre-Existing			-0.417**	
			(0.181)	
Games × Post × Small Type				-1.244***
				(0.158)
Observations	70	1,470	1,470	980
R-squared	0.905	0.667	0.669	0.624
Unit of Observation	Game/Non-Game	App-Type	App-Type	App-Type
Time Period	Jan 12 / Dec 14	Jan 12 / Dec 14	Jan 12 / Dec 14	Jan 12 / Dec 14
Sample	All	All	All	All Non-Games + Action, Arcade Card and Casino
Year/Month FE	•	•	•	•
App-Type FE		•		•

Notes: Sample period in all columns and panels is January 2012 to December 2014. Sample in Column (1) includes monthly observations at the Game/Non-Game level. Sample in Columns (2) and (3) includes all game and non-game observations at the app-type level. Sample in Column (4) includes all non-game and Action, Arcade, Card and Casino app-type observations at the monthly level. Outcomes in panel (a) are the absolute number of new entrants at the game/non-game or app-type level. Outcomes in panel (b) are the average share of 1-star ratings at the game/non-game or app-type level. Outcomes in panel (c) are mean app size as expressed by the natural log of size in MB. Outcomes in panel (d) the average ratings of entrants at the game/non-game or app-type level, where each entrant's weight is determined by their downloads (rather than a simple mean). Controls include year/month fixed effects, a "Game" category group dummy for odd columns and app type fixed effects. Additional controls include game/non-game or app-type specific time trends. The variable "Games × Post" is equal to 1 for games (or game types for even columns) during and after March 2014 and zero otherwise. Standard errors are robust to heteroskedasticity in Column (1) and clustered at the app-type level otherwise. *** p<0.01, ** p<0.05, * p<0.1

C.7 Entry and Quality: Alternative Outcomes

Table C9 combines the main results from Tables 5 and 6 in the main text but includes alternative outcome variables. Panel (a) uses *absolute* entry numbers. These regressions confirm the results from the log-transformed estimates in the main text. Absolute entry for the average app-type increases by about 1,500 apps, and for all games by almost 35,000 apps after re-categorization. As in the log estimates in Table 5, these are large treatment effects, given the size of the average game app-type. Columns (3) and (4) show that app-types where discovery costs fall by more in response to re-categorization are the ones driving the effects.

Panels (b) and (c) test alternative proxies for quality. Panel (b) uses the average share of 1 star ratings received by entrants and panel (c) uses app size in ln(MB). This measure could capture quality since additional app features (i.e., 3D graphics, databases, video chatting) require more lines of code which increases size. Estimates largely confirm findings in the main text, showing a decrease in the quality of new games following re-categorization. In both cases, at the game/non-game level, difference-in-differences coefficients are not statistically significant. At the app-type level, coefficients are statistically significant at the 95% confidence level. The share of 1 star ratings for game apps increases (consistent with the decrease in average ratings) and average game app size falls relative to non-games. Heterogeneity in these effects across different game app-types are also similar to the main text, except for Column (3) in panel (c) which does not find statistically significant differences in app size between games with an without pre-existing categories after re-categorization. This lack of heterogeneity may reflect the roughness of this particular proxy for app quality.

Quality outcomes in Tables 5 and 6 are based on simple averages. This may not fully capture the *realized* quality that consumers experience in the market. To account for this, I calculate download-weighted quality measures: average ratings weighted by app downloads. Estimates from regressions with these outcomes are in panel (d) of Table C9. Coefficient estimates are qualitatively the same as previous results and suggest that quality effects are not driven by the simple averaging of app quality. Coefficient estimates in the weighted outcome regressions are larger than in the simple average regressions. Weighted average game app ratings decrease by 0.5 stars, on average. This means that the effects of re-categorization on quality are likely *understated* in the main text.

C.8 Timing Tests

I allow treatment effects to vary over time by introducing interactions between monthly date dummies and the treatment group dummy. I estimate timing tests both at the aggregate game/non-game level and at the less aggregate app-type level. The estimating equation, for app-type c at time t is:

$$y_{ct} = \sum_{t=\text{re-cat. month}-10}^{\text{re-cat. month}+10} \tau_t \left(Game_c \times D_t\right) + \delta_c + \delta_t + \epsilon_{ct} \tag{2}$$

where y_{ct} is an outcome and where τ_ts now capture period specific treatment effects relative to a baseline period. D_t is a dummy equal to one for observations during month t and zero otherwise. Since I have 10 periods after re-categorization I test the 10 periods before re-categorization for parallel pre-trends relative to the time before April 2013. As before, I include game/non-game, app-type and time fixed effects, and game/non-game or app-type specific time trends. Figure C3 shows the period specific treatment effects for the main measures of entry and quality used in the main text. For each outcome, game/non-game level results are on the left hand panel, and app-type results are on the right hand panel.

Entry estimates show that treatment effects become statistically significantly different from zero exactly around re-categorization. There are no treatment effect estimates which are statistically significantly different than zero (at the 95 percent confidence level) in the 10 periods before February 2014. February 2014 itself (two months following the announcement) has a statistically significant positive coefficient, likely representing a response by developers to the announcement of new categories in December 2013. Some apps may have entered the market early to position themselves in anticipation of the change.[12] Entry response happens quickly after the announcement since apps have short development time (see Section 2.2 in the main text). Developers can create simple apps in as little as a month.[13] Point estimates are highest right after re-categorization takes place.

The timing test for quality looks qualitatively similar to entry although estimates are noisier. At the app-type level, monthly treatment effects are statistically significant at the 95 percent confidence level following the announcement.[14] Treatment

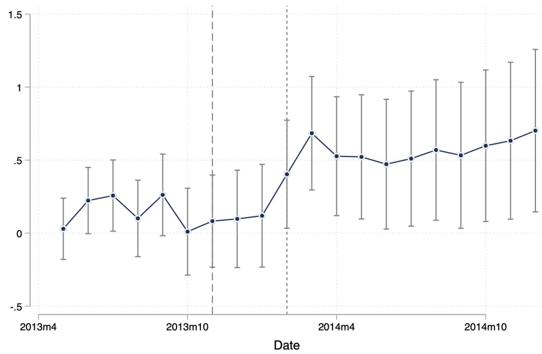
[12] The announcement did not set a strict date for the implementation of new categories, but said that the change will happen in the first quarter of 2014 (9to5Google.com)

[13] New entry could have come from multiple sources. Developers creating completely new apps, porting existing apps from the iOS store, or releasing already developed products into the market early.

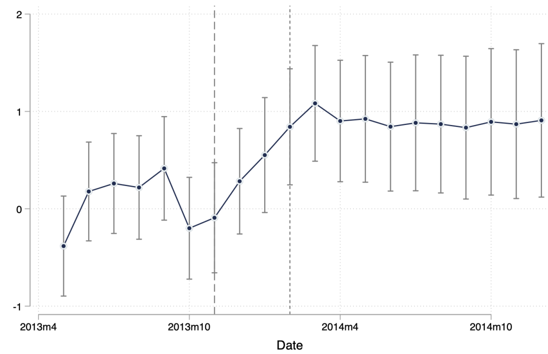
[14] In the game/non-game panel, estimates are too noisy to be statistically significant, but coeffi-

Figure C3: Entry Timing Tests

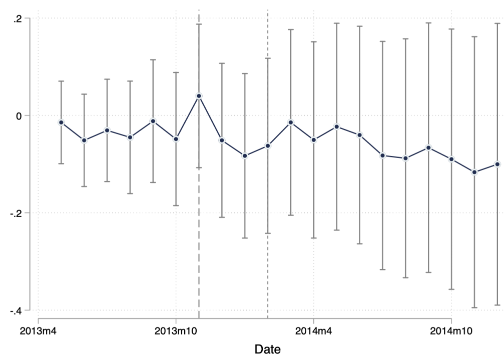
(a) Outcome: ln(N Entrants)



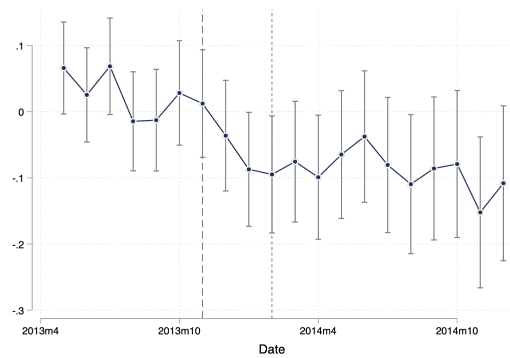
(b) Outcome: ln(N Type Entrants)



(c) Outcome: Mean Entrant Rating



(d) Outcome: Mean Type Entrant Rating



Notes: Each panel shows estimates of coefficients τ_t from Equation 2 at different aggregation levels and with different outcomes. Panels (a) and (c) are estimated at the game/non-game level. Panels (b) and (d) are estimated at the app-type level. Data from January 2012 to December 2014 is used throughout. Additional controls in each regression include year/month fixed effects, game/non-game or app-type fixed effects, and game/non-game or app-type specific trends. Standard errors for panels (a) and (c) are robust to heteroskedasticity and standard errors for panels (b) and (d) are clustered at the app-type level. 95% confidence intervals shown. In each panel, the first dashed vertical line represents the announcement of re-categorization and the second dashed vertical line represents the start of the re-categorization period.

effects before the announcement are statistically zero but entrant quality falls between the announcement and the actual re-categorization. This mitigates a concern that rating-based quality reflects "subjective" rather than "objective" quality. Rat-

cient estimates point in the same direction.

ings measures would be "subjective" if they depend on comparisons that consumers make within groups of similar apps. This would be a problem since comparison groups likely change following re-categorization. Observing effects ahead of physical changes in the store (but after the announcement) suggests that ratings are not "subjective." After re-categorization, effects are persistent until the last sample month. It may be reasonable to expect quality effects to disappear over time if low-quality apps require little time to program. Under this mechanism additional low quality apps would enter first and reduce average quality, but higher quality apps would eventually enter. Fishman and Levy (2015) and Moraga-González and Sun (2020) suggest that a change in search costs could *permanently* reduce the incentives of firms to provide quality in equilibrium. My estimates show a relatively long-term quality change in the app market.

The timing tests also help address a concern with the overall findings: that changes I find are due to changes in consumer preferences rather than discovery costs. This is an unlikely explanation since changes in entry and quality anticipate the actual re-categorization. If the effect of re-categorization was on consumer preferences, it would take time for firms to learn about these changes and respond. Changes in consumer discovery costs, given knowledge of future categories, are easily anticipated by firms.

C.9 Quality Change Mechanisms

There are two possible mechanisms that can explain why entrant quality falls after a change in discovery costs. In the first mechanism, as discovery costs fall, consumers search the store more and are more likely to download apps. This is analogous to a market expansion in the app market. Some apps that did not receive enough downloads to cover entry and operational costs under high discovery costs find it profitable to enter into the market under low discovery costs. A two-stage investment and entry model such as Melitz (2003) predicts that these effects should be more pronounced for infra-marginal lower quality apps and more lower quality apps should enter the market and reduce average entrant quality.

Fishman and Levy (2015) propose an alternative mechanism for lower discovery costs reducing average entrant quality. In this mechanism, consumers have horizontal and vertical preference heterogeneity. In a high discovery cost world, consumers see few products, some high quality and some low quality. Returns to quality are high as the main comparison consumers make is vertical. In a low discovery cost world, consumers see many high and low quality products and make both vertical and horizontal comparisons. If horizontal preference heterogeneity is strong enough, developers realize that investment in high quality products does not matter as much, since consumers will simply pick the product that best fits their horizontal taste. Investment in quality falls.

Results in the main text show that app-types with greater decreases in search costs have more entry and more lower quality entry. This is consistent with additional infra-marginal entry. The Fishman and Levy (2015) mechanism predicts that quality reductions in response to discovery cost decreases should be greater for higher quality products. I test whether this is the case by examining changes in different moments of the new entrant rating distribution. I calculate the 5th, 25th, 50th, 75th and 95th new entrant rating percentiles for each game and non-game app-type and month. I then compare ratings at each percentile in each time period before and after re-categorization for game app-types relative to non-game app-types. For example, I compare what is the rating at the 25th percentile of the new entrant rating distribution for game app-types as compared to non-game app types. If the Fishman and Levy (2015) mechanism is present, average quality should decrease at each percentile, and rating decreases at higher percentiles should be bigger than rating decreases at lower percentiles. If it is not present and quality falls primarily through additional entry of infra-marginal apps, ratings should decrease more at lower percentiles and not change much at higher percentiles, since entry incentives for high quality apps do not change substantially.

Estimates from this set of regressions are in Table C10. They show that ratings fall more at the lower end of the quality distribution than at the higher end for games after re-categorization. The 5th percentile of the entrant rating distribution decreases by nearly 0.2 stars and the 25th percentile of the entrant rating distribution falls by 0.1 stars. By comparison, the 50th and 75th percentiles of the entrant rating distribution do not change after re-categorization. Average ratings at the 95th percentile of the distribution fall, but only by about 0.04 stars. Overall, these results suggest that re-categorization primarily increased entry among infra-marginal apps that were previously not profitable and did not change entry much at the top of the quality distribution.

Even though quality at the top of the entrant distribution does not appear to change, increases in low quality entry can still have substantial welfare impact on consumers. With congestion externalities, additional low quality apps can limit consumer discovery of high quality apps. I decompose the welfare effects of re-categorization in Section 5.3.

Table C10: Difference in Differences: Quality Mechanism

Outcome:	Entrant Avg. Rating at 5th Percentile of App-Type Ent. Rating Dist.	Entrant Avg. Rating at 25th Percentile of App-Type Ent. Rating Dist.	Entrant Avg. Rating at 50th Percentile of App-Type Ent. Rating Dist.	Entrant Avg. Rating at 75th Percentile of App-Type Ent. Rating Dist.	Entrant Avg. Rating at 95th Percentile of App-Type Ent. Rating Dist.
Games × Post	-0.162***	-0.097***	-0.041	0.003	-0.038***
	(0.052)	(0.031)	(0.029)	(0.023)	(0.012)
Observations	1,469	1,469	1,469	1,469	1,469
R-squared	0.551	0.575	0.745	0.739	0.437
Unit of Observation:	App-Type	App-Type	App-Type	App-Type	App-Type
Sample Period:	Jan 2012/Dec 2014	Jan 2012/Dec 2014	Jan 2012/Dec 2014	Jan 2012/Dec 2014	Jan 2012/Dec 2014
Sample:	All	All	All	All	All
Year/Month FE	•	•	•	•	•
App-Type FE	•	•	•	•	•

Notes: Sample period in all columns covers January 2012-December 2014. Sample in all columns consists of monthly observations at the app-type level. Outcomes are the average rating of new entrants at the Xth percentile of the new entrant rating distribution, where X is the 5th, 25th, 50th, 75th, or 95th percentile, depending on the column. Controls include year and month fixed effects, app-type fixed effects and app-type time trends. The variable "Games × Post" is a dummy variable equal to 1 for games, or game app-types during and after March 2014. Standard errors are clustered at the app-type level. *** p<0.01, ** p<0.05, * p<0.1

C.10 Prices

Figure C4 plots three graphs showing price patterns in the Google Play Store. Panel (a) shows the ratio of mean paid game prices over mean paid non-game prices. Panel (b) shows the ratio of mean *new* paid game prices over mean *new* paid non-game prices. Prices do not appear to change substantially.

Panel (a) shows average prices for all games falling as compared to non-games, potentially due to increasing importance of in-app advertising and in-app purchases in the app economy. After re-categorization, average game prices increase and the ratio of game to non-game prices stabilizes. Lower discovery costs from the re-categorization could be the cause of the price changes.[15] However, in absolute terms, the magnitudes of changes are small. In panel (b), it is apparent than there are no substantial differences in the prices of new games relative to non-games. The price ratio before and after is similar on average.

I estimate difference-in-differences regressions with both average prices and average entrant prices as outcome variables. Results for these regressions are in Table C11 at the game/non-game and at the app type. They show that there are no statistically significant differences between game and non-game paid app prices after re-categorization as compared to before. There is also no statistically significant heterogeneity (at the 95% confidence level) across game app-types that were more or less affected by re-categorization. This is true regardless of whether I look at all paid apps in panel (a) or only at new paid apps in panel (b).

Panel (c) of Figure C4 shows the ratio of the share of new paid games appearing

[15] With lower costs, higher valuation consumers can discover more preferred game-apps more easily (Bar-Isaac et al. 2012).

in a given month (as a percentage of the total number of new games), over the share of new paid non-games (as a percentage of the total number of new non-games). Changes in the revenue streams of paid and non-paid apps (e.g., the increasing prevalence of in-app purchases) may result in changes in the number of entrants into the market. Such changes could also drive app entry and undermine the search mechanism explanation. This does not appear to be the case in the data. Panel (c) shows that there are no changes in the patterns of free and paid product entry between games and non-games after re-categorization.[16]

In addition to the difference-in-differences estimates, I also test for whether changes in the number of other apps in a category affect a paid app's prices. As in Columns (7) and (8) of Table 4 I use short run changes in the number of apps in game categories due to re-categorization. Apps move from being in large categories with many other apps of different types to smaller categories with other apps of their own type. One explanation for the findings in the main text showing that downloads for games with more apps in their category fall is that these apps face more competition from imperfect substitutes. If this is the case, there should also be a link between the number of apps in a category and prices.

In Table C12 I show the results of a regression relating pre/post re-categorization differences in the number of apps in game categories to pre/post differences in individual app prices. Coefficient estimates are both small in absolute terms and are statistically null. There is no relationship between changes in the number of apps in a category and changes in app prices. This suggests that changes in the number of apps in a category does not affect competition. Instead, it primarily affects the market by reducing congestion.

C.11 Google Trends Evidence of Consumer Awareness of Android Games and Non-Games

I do not observe Google's advertising for the Google Play app store. Instead, I use Google Trends search volumes to proxy consumer awareness for Android Games and Android Apps. Figure C5 shows the weekly Google Trends volumes from January 2012 to December 2015. The top two panels compare Google Trends for the "Android Games" and "Android Apps" search queries. The middle panels compare "Google Play Games" and "Google Play Apps." The last two panels compare "Google Play Games" and "iOS Games." In all cases Google trends are measured

[16] There are substantial changes in the absolute share of paid apps that are entering into the market over time. The share of new paid products falls from over 30% to less than 10%. This pattern is consistent for both games and for non-game apps.

Table C11: Prices: Difference-in-Differences Estimates

Panel (a): All Paid Apps

Outcome Variable:	Mean Price	Mean Price	Mean Price	Mean Price
Games × Post	0.037	-0.037	-0.147	0.019
	(0.036)	(0.122)	(0.152)	(0.105)
Games × Post × No Pre-Existing			0.197	
			(0.138)	
Games × Post × Small Type				-0.595*
				(0.347)
Games	9.966***			
	(0.758)			
Observations	70	1,470	1,470	980
R-squared	0.999	0.970	0.970	0.966

Panel (b): Paid Entrant Apps

Outcome Variable:	Mean Price	Mean Price	Mean Price	Mean Price
Games × Post	0.379	0.076	0.030	0.000
	(0.443)	(0.423)	(0.470)	(0.459)
Games × Post × No Pre-Existing			0.082	
			(0.288)	
Games × Post × Small Type				-0.050
				(0.910)
Games	28.088**			
	(10.409)			
Observations	70	1,470	1,470	980
R-squared	0.999	0.970	0.970	0.966
Unit of Observation:	Game/Non-Game	App-Type	App-Type	App-Type
Sample:	All Paid	All Paid	All Paid	All Paid
Sample Period:	Jan 12/Dec 14	Jan 12/Dec 14	Jan 12/Dec 14	Jan 12/Dec 14
Year/Month FE	•	•	•	•
App-Type FE		•	•	•

Notes: Sample period in all columns covers January 2012-December 2014. Sample in Column (1) consists of monthly observations at the Game/Non-Game level. Sample in Columns (2)-(4) consists of monthly observations at the app-type level. Outcomes in panel (a) are average prices for all paid apps at each aggregation level. Outcomes in panel (b) are average prices for all paid new entrants at each aggregation level. Controls include year and month fixed effects, and game/non-game fixed effects or app-type fixed effects, depending on the column. Additional controls include game/non-game or app-type time trends. The variable "Games × Post" is a dummy variable equal to 1 for games, or game app-types during and after March 2014. Standard errors are robust to heteroskedasticity in Column (1) and are clustered at the app-type level in the remaining columns. *** p<0.01, ** p<0.05, * p<0.1

relative to the maximum search volume over the period. The figures on the left are absolute search trends numbers and the figures on the right are search trend ratios. Google Play/Android Games volumes are always the numerators in the ratios.

The figures all show that there is substantial variation in search query volumes over the sample period. For example, there is a spike in search queries around Christmas/the New Year. In all three sets of comparisons there is no spike in Google Play/Android search queries around the period of the re-categorization of the store

Table C12: Price Changes in Response to Changes in Number of Apps in a Category

	(1) Post/Pre Δ ln(Price)	(2) Post/Pre Δ Price
Post/Pre Δ ln(N Apps in Category)	-0.000	-0.005
	(0.002)	(0.009)
Observations	21,749	21,749
R-squared	0.449	0.272
Unit of Observation:	App	App
Sample Period:	Jan 14 / Apr 14	Jan 14 / Apr 14
Sample:	All Paid Games	All Paid Games
App Controls	•	•

Notes: Sample period in both columns covers January 2014 to April 2014. Sample includes monthly observations of all paid game apps present from January 2014 to April 2014. Additional app-level controls include average app ratings and app age-specific fixed effects. The outcomes are differences between app-level average price in March and April 2014 and app-level average price in January and February 2014. Δ ln(N Apps in Category) is the difference in the natural log of the number of apps in the category of app j after re-categorization (March and April 2014) and the natural log of the number of apps in the category of app j before re-categorization (January and February 2014). Standard errors are robust to heteroskedasticity. *** p<0.01, ** p<0.05, * p<0.1

(solid vertical red line). There is also no change in the relative search query ratio around the period of the re-categorization. In the first two panels (Android Games vs. Android Apps), the "Google Play Games" search query is trending upwards relative to the "Google Play Apps" search query. There is no change in this trend around the re-categorization period.

C.12 Developer Switching Between Games and Non-Games

Figure C6 shows a ratio of the number of existing non-game developers[17] who produce a game in period t over the number of existing game developers who produce a non-game in period t. The ratio is greater than 1: there are more non-game developers switching to producing games than the other way around. The figure also shows that there is an increase in the ratio around the period of the game re-categorization (from 2 to 2.6). This is potentially consistent with a resource allocation story whereby developers have a fixed budget and have to choose between producing games and non-games. However, within 2 months of the re-categorization, the ratio falls to pre re-categorization levels.

This suggests that developer switching after re-categorization is a short term response. By comparison, the entry effects of re-categorization are a long term phenomenon. Period-specific treatment effects captured in Figure C3 show that the increase in the number of games relative to non-games persists all the way to the end

[17] Defined as those who only developed non-games in the past.

of the sample (9 months after re-categorization). The magnitude of the treatment coefficient in the last month of the sample in Figure C3 is as large as the magnitude in the second month after re-categorization. This suggests that the magnitude of the treatment effect cannot be explained by developers switching from producing non-games to producing games.

D Appendix D

D.1 Demand Model with Search

This section describes a demand model with search following the consideration set approach of Moraga-González et al. (2015).[18] Consumers choose a single product out of a set of N products. For each product j, consumers obtain utility $u_{ij} = \delta_j + \epsilon_{ij}$. Consumers are not fully informed about products: they do not know their ϵ_{ij}s. Search resolves this uncertainty. Consumers in this market first choose a consideration set A of products and pay a set-specific search cost. They find out the ϵs of those products and pick product j out of subset A. In this application, the products in subset A can be located across multiple categories. Subsets are unobserved to the econometrician. Consumers know the expected utility (or inclusive value) they obtain from the products in subset A: $\overline{U}_A$.[19] Consumers incur subset-specific search costs (c_{iA}) such that the utility of consumer i choosing subset A is:

$$u_{iA} = \overline{U}_A - c_{iA} = \overline{U}_A - \left(\sum_{r \in A} \gamma \psi_r + \lambda_{iA} \right) \tag{3}$$

where ψ_r reflects a deterministic "distance" between the consumer and product r in set A. λ_{iA} is a consumer/choice set specific search cost shock, which I assume is EV type 1 distributed mean zero with a standard error normalized to 1.[20] This shock can be interpreted as an information shock - word of mouth from friends or family. γ is effectively the average marginal search cost for consumers in the market. As with consumer utility, search costs have no unobservable heterogeneity aside from the idiosyncratic shock.

Due to the idiosyncratic error terms on both search costs and consumer utility, the unconditional probability of a consumer choosing product j is:

[18] It is also similar to Goeree (2008) and Honka et al. (2017).

[19] In a multinomial logit model, this is simply $log[1 + \sum_{r \in A} exp(\delta_r)]$. Consumers also always have the outside option, regardless of the set they consider.

[20] γ can vary across products or product groups.

$$P_j = \sum_{A \in A_j} P_A P_{j|A} \tag{4}$$

where A_j is the set of all subsets that product j belongs to, P_A is the probability of a consumer choosing subset A From the set of all possible subsets and $P_{j|A}$ is the probability that consumer i picks product j out of subset A. The unconditional probability P_j is equivalent to the observed market share of product j (s_j). Moraga-González et al. (2015) show that it is possible to "integrate out" the unobservable subsets and obtain the following closed form expression for s_j:[21]

$$s_j = \frac{\frac{exp(\delta_j)}{1+exp(\gamma\psi_j)}}{1 + \sum_{k \in N} \frac{exp(\delta_k)}{1+exp(\gamma\psi_k)}} \tag{5}$$

where the denominator sums up over all products in the market (N) rather than over specific subsets. This expression is effectively the standard multinomial logit model except that the market share of product j is shaded down by how hard it is to find (ψ_j). It is similar to the market share specification in Equation 6 in the main text. Setting $exp(R_{jc}) = \frac{1}{1+exp(\gamma\psi_j)}$ and introducing an additional nested logit error term equates the two.[22]

I estimate this model using non-linear GMM with the same instruments used to estimate the linear demand model in the main text. Demand estimates from this model are in Column (3) of Table D2. These are qualitatively similar to demand estimates from the main text.

While this is a reasonable approach to modelling consumer product discovery and demand in the mobile app market, there are potentially many other ways in which consumers search the market. This model also does not easily allow to control for additional unobservable heterogenity with aggregate product level data. I choose to use the simpler linear demand model in the main text. It does not make specific assumptions about the consumer search process but is broadly consistent with many predictions from search literature.

[21] The assumption that the "distance" of products in a consideration set is additive in the set's search costs is key for obtaining a closed form expression for choice probabilities.

[22] This consideration-set based model allows for unobservable heterogeneity in consumer preferences such as a consumer/category specific shock. However, the standard market-share inversion procedure for nested logit models does not apply to the consideration set model and it would have to be estimated by simulation. I do not include additional unobservable heterogeneity for this reason.

D.2 Additional Demand Model Regressions

Table D1: **1st Stage Supporting Regression**

	(1)
Outcome Variable:	ln(N Category Apps$_t$)
ln(N Category Apps$_{t-1}$)	0.142***
	(0.022)
Mean Category Rating$_t$	-0.279***
	(0.065)
ln(Category Downloads$_t$)	0.387***
	(0.034)
Category FE	•
Observations	624
R-squared	0.953

Notes: The sample includes monthly category-level observations from February 2012 to December 2014. Standard errors are robust to heteroskedasticity. *** p<0.01, ** p<0.05, * p<0.1.

D.3 Placebo Time-Varying Fixed Effects

In the main text, I include two sets of time-varying fixed app-type fixed effects: a set of app-type fixed effects that turns on before re-categorization takes place, and a set of app-type fixed effects that turns on after re-categorization takes place. The difference in these fixed effects is in Figure 3 and shows app-type level welfare changes. I interpret these changes as being driven by re-categorization improving information, but they could also be caused by other time varying heterogeneity within app-types. For example, Educational games have the biggest fixed effect change, which could be the result of consumers liking educational games more over time.

To test whether the variation in fixed effects is driven by the re-categorization event, I introduce a specification of the model with three sets of time-varying app-type fixed effects. The first set of app-type fixed effects is active from March 2012 to February 2014. The second set of app-type fixed effects is active only during April 2014 (March 2014 is omitted from the data) and the last set of app-type fixed effects is active from May 2014 to December 2014. The change between the first two sets identifies information effects just around re-categorization. The change between the second two sets identifies whether there were other changes over time. If changes in app-type fixed effects primarily capture changes in category informativeness, I should not see substantial differences between the April 2014 and May-December 2014 fixed effects where informativeness was constant.

Table D2: **Additional Demand Estimates**

	(1)	(2)	(3)
ln(N Apps in Category)		-0.392***	0.374***
		(0.026)	(0.016)
ln(N Apps in Category) × New App		-0.012	
		(0.011)	
ln(Lag App Downloads)		0.036***	-0.282***
		(0.004)	(0.008)
σ	1.404***	0.708***	
	(0.069)	(0.027)	
Price	-3.800***	-0.833***	-1.470***
	(0.119)	(0.111)	(0.041)
ln(Size in MB)	0.078***	0.046***	0.099***
	(0.007)	(0.006)	(0.002)
Video Preview Dummy	-0.016	0.072***	0.138***
	(0.020)	(0.017)	(0.005)
N Screenshots	0.000	0.008***	0.022***
	(0.002)	(0.001)	(0.000)
Paid and New App Dummies	•	•	•
App Rating FE	•	•	•
Year/Month FE	•	•	•
Category FE	•	•	•
Developer FE	•	•	•
Observations	4,152,147	4,152,147	4,167,060

Notes: The sample includes monthly observations of all free and paid game apps in the Google Play Store from March 2012 to December 2014, excluding March 2014. Column (1) shows estimates of a nested logit model without discovery friction controls. Column (2) shows estimates of the model from the main text with heterogeneous discovery frictions for new and incumbent apps. Column (3) shows estimates of the search and demand model described in Online Appendix D.1. "App Rating FE" are a set of dummies representing the average rating of app j in period t within 0.5 stars. Apps with 2 stars or less are the "baseline" category for "App Rating FE." $Year/MonthFE$ include year and month dummies. Instruments for price and for σ include the ratings of other apps in the same category, the number of screenshots of other apps of the same app-type and the average size of other apps of the same app-type. Instruments for lagged downloads for app j include differences in the lags of app j downloads 2 and 3 periods before period t. Instrument for the number of apps in the category is described in the main text and is the residual of the regression in Table D1. Standard errors are clustered at the app level in Columns (1) and (2) and robust to heteroskedasticity in Column (3). *** p<0.01, ** p<0.05, * p<0.1.

Figure D1 shows both sets of differences for each app-type and their computed 95% confidence intervals. The results suggest that consumer utility from app types changed around the period of re-categorization and not after. Most of the "placebo" fixed effect differences are statistically zero at the 95% confidence level. Even for the few app-types where these differences are not statistically zero, they are very small in magnitude relative to the true difference in fixed effects between the pre- and post-re-categorization.

D.4 Distributions of Main Model Welfare Effects

Figure D2 shows the full distribution of outcomes generated by the randomized simulations. The figures show that the randomizations do not substantially change the main effects. For congestion effects, the values from different randomizations are clustered tightly around the mean. For entrant quality effects, the vast majority of randomizations are within 15% (0.001) of the mean effect.

Online Appendix References

Bar-Isaac, H., G. Caruana, and V. Cuñat Martinez (2012). Search, design and market structure. *American Economic Review 102*(2), 1140–1160.

Chevalier, J. A. and D. Mayzlin (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research 43*(3), 345–354.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review 94*(5), 1429–1451.

Garg, R. and R. Telang (2013). Inferring app demand from publicly available data. *MIS Quarterly 37*(4), 1253–1264.

Goeree, M. S. (2008). Limited information and advertising in the us personal computer industry. *Econometrica 76*(5), 1017–1074.

Honka, E., A. Hortaçsu, and M. A. Vitorino (2017). Advertising, consumer awareness, and choice: Evidence from the us banking industry. *The RAND Journal of Economics 48*(3), 611–646.

Kummer, M. and P. Schulte (2019). When private information settles the bill: Money and privacy in google's market for smartphone applications. *Management Science 65*(8), 3470–3494.

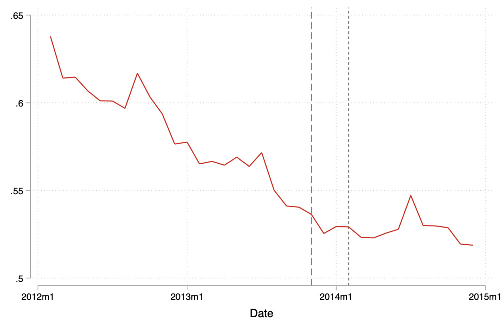
Leyden, B. T. (2018). There's an app (update) for that. mimeo.

Liu, Y., D. Nekipelov, and M. Park (2014). Timely versus quality innovation: The case of mobile applications on itunes and google play. *NBER Working Paper*.

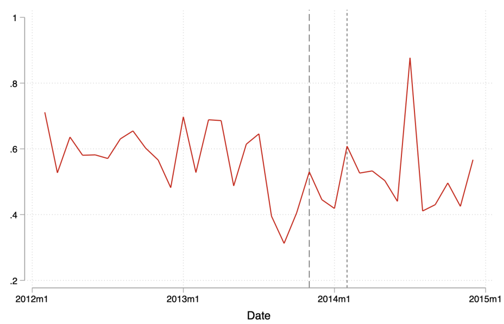
Moraga-González, J. L., Z. Sándor, and M. R. Wildenbeest (2015). Consumer search and prices in the automobile market.

Figure C4: Prices

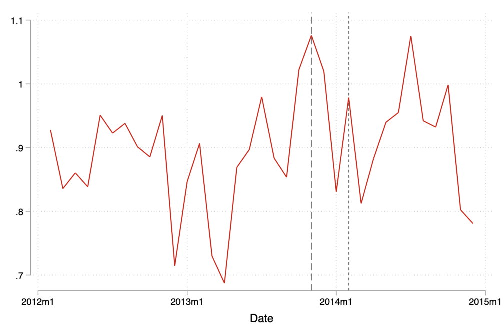
(a) Outcome: Mean Game Price / Mean Non-Game Price



(b) Outcome: Mean Entrant Game Price / Mean Entrant Non Game Price



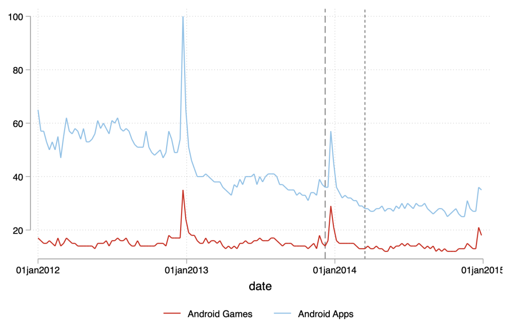
(c) Outcome: Share Paid Game Entrants / Share Paid Non-Game Entrants



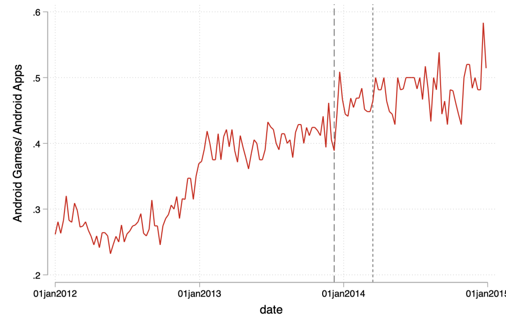
Notes: Panel (a) shows a ratio of mean monthly game app price over mean monthly non-game app price using all paid apps. Panel (b) shows a ratio of mean monthly game app price over mean monthly non-game apps price using only paid entrants. Panel (c) shows a ratio of the monthly percentage of new game apps that are paid over the monthly percentage of new non-game apps that are paid. In all panels, the first dashed vertical line represents the re-categorization announcement and the second dashed vertical line represents the start of the re-categorization period.

Figure C5: US Google Search Trends

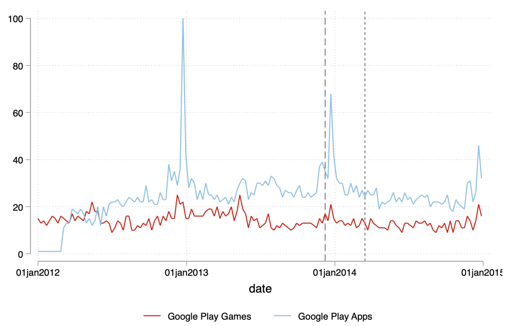
(a) “Android” Searches



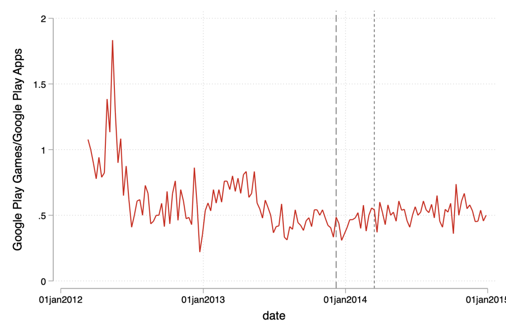
(b) Ratio of Panel (a) Games/Apps



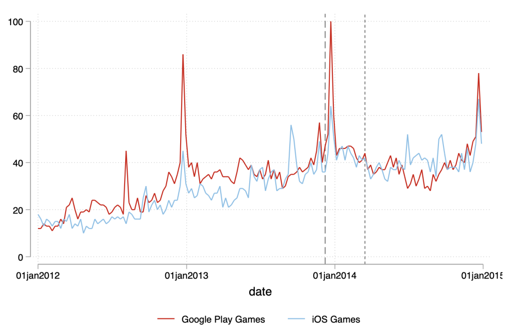
(c) “Google Play” Searches



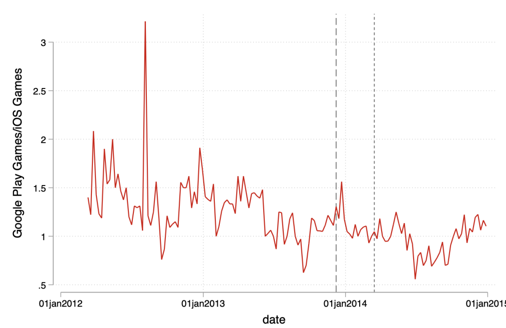
(d) Ratio of Panel (c) Games/Apps



(e) Google Play/iOS Searches

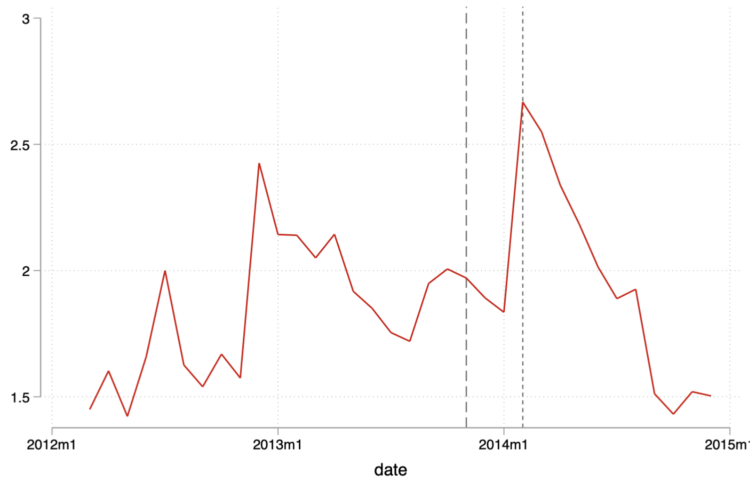


(f) Ratio of Panel (e) Google Play/iOS



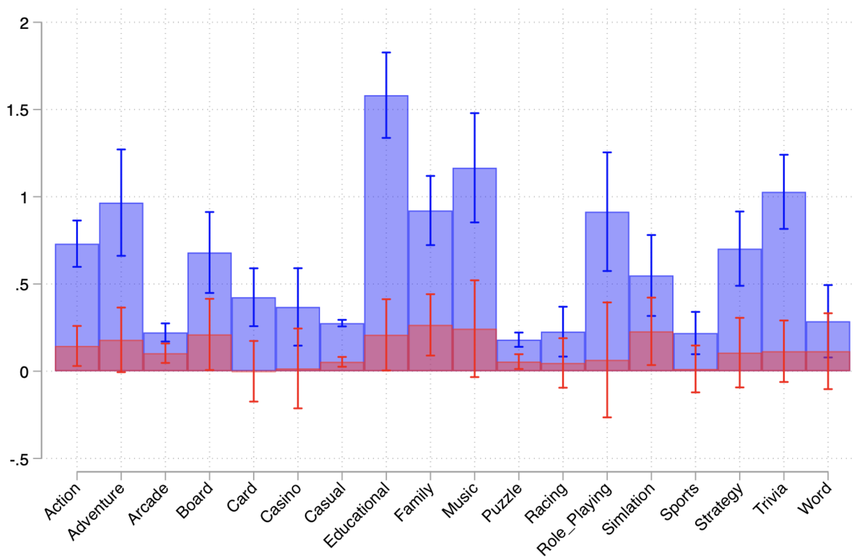
Notes: Panels (a), (c) and (e) show daily Google Trend search volume estimates for different queries. In each of the panels, numbers are normalized relative to maximum search volume which is set to 100. Panels (b), (d) and (f) show ratios of the numbers in panels (a), (c) and (e), respectively. In all panels, the first dashed vertical line represents the re-categorization announcement and the second dashed vertical line represents the start of the re-categorization period.

Figure C6: Ratio of Switching Developers: $\frac{\text{Non-Game to Game}}{\text{Game to Non-Game}}$



Notes: This figure shows a monthly ratio. In each month t the numerator is the number of developers who produced a non-game app in any period before t and produced a game app in period t. The denominator is the number of developers who produced a game app in any period before t and produced a non-game app in period t. The first dashed vertical line represents the re-categorization announcement and the second dashed vertical line represents the start of the re-categorization period

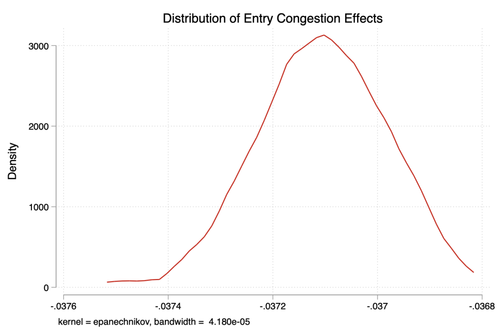
Figure D1: Differences in App-Type Fixed Effects



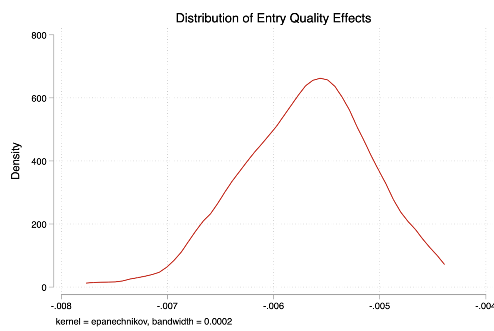
Notes: Each column shows two sets of differences in estimated app-type fixed effects based on the model described in Section 5.1. The first difference, in blue, is between the app-type fixed effect for Mar 2012-Feb 2014 and the app-type fixed effect for April 2014 $\gamma^1_{c,Apr14} - \gamma^1_{c,Mar12-Feb14}$. The second difference, in red, is between the app-type fixed effect for April 2014 and the app-type fixed effect for May 2014-December 2014. 95% calculated confidence intervals for both differences are shown.

Figure D2: Distribution of Welfare Effects Across Simulations

(a) Entry Congestion Effects (mean = -0.037)



(b) Entrant Quality Effects (mean = 0.006)



Notes: Panel (a) shows the full distribution of entry congestion effects for 100 simulations. As in Table 8, for each simulation this is calculated as the difference $EU_4 - EU_7$, where both EU_4 and EU_7 are defined in Section 5.3. Panel (b) shows the full distribution of of entrant quality effects for 100 simulations. As in Table 8, for each simulation this is calculated as $EU_5 - EU_6$, where both EU_5 and EU_6 are defined in Section 5.3.