Variety-Based Congestion in Online Markets: Evidence from Mobile Apps

By Daniel Ershov*

In many online markets, consumers have to spend time and effort browsing through products. The addition of new products could make other products less visible, creating congestion externalities. Using Android app store data, I take advantage of a natural experiment – a re-design of part of the store – to show evidence of congestion externalities online: more apps in the market directly reduce per-app usage/ downloads. The natural experiment also increases long-run entry, but a structural demand model that accounts for congestion externalities suggests that forty percent of consumer variety welfare gains are lost from higher congestion.
JEL: D12, D22, L86
Keywords: congestion, mobile apps, online competition, entry

In the last 30 years, digital-technology related reductions in production, distribution, and transportation costs facilitated the creation of many new products. Information-aggregating online marketplaces / platforms unlocked consumer access to these products. However, actually discovering new products online remains a challenge for consumers. Despite platform features such as personalized recommendations, search algorithms, product categories, bestseller lists, and ratings, consumers have to spend substantial time and effort examining products. Limited attention and screen size creates congestion, as consumers may not see some products because of the existence of others. Such congestion is a source of consumer and firm concerns in virtually every online market (Mull 2019), and could undercut the benefits of growing variety.

This paper presents the first empirical evidence on the existence and magnitude of congestion effects online. I use data on the Google Play (Android) mobile app store from January 2012 to December 2014. This is 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 (Bruemmer 2012). Consumers primarily discover products by browsing through

* Ershov: UCL School of Management, 1 Canada Square, E14 5AA, London, d.ershov@ucl.ac.uk. I am grateful to the editor, Leslie Marx, and three anonymous referees for their helpful comments. This paper was a chapter in my PhD dissertation and previously circulated under the titles "The Effects of Consumer Search Costs on Entry and Quality in the Mobile App Market," and "Consumer Product Discovery Costs, Entry, Quality and Congestion in Online Markets." I would like to thank Victor Aguirregabiria, Avi Goldfarb and Heski Bar-Isaac. This paper also substantially benefitted from discussions with Joel Waldfogel, Eduardo Souza Rodrigues, Mar Reguant, Sandro Shelegia, and comments from participants at numerous seminars and conferences. 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.

platform-defined categories (e.g., "Productivity Apps").[1] Many other online markets such as eBay, Amazon, and Netflix are similarly organized.

To establish the existence of congestion externalities, 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-organized its 6 game categories into 18 categories (Table 1), reducing the number of apps per category.[2] As non-game categories were not changed, I compare game and non-game outcomes using a difference-in-differences approach. I show that, in the very short-run, downloads increased for games relative to non-games, and most importantly, downloads increased by more for apps that ended up in less populous categories. This points to the presence of congestion externalities, as apps in categories with fewer other apps are more likely to be visible to consumers.

The re-design also had an effect on the supply of game apps, increasing game entry by approximately 34%. Although greater variety should increase welfare, the existence of congestion externalities should mitigate welfare gains from entry. Drawing on a consideration-set framework, I estimate the parameters of an app demand model that accounts for congestion externalities. I then evaluate the welfare effects of congestion. Each consumer gained \$0.057 per-month from the additional product variety in the market after re-categorization. This adds up to nearly \$70 million per-year across all US Google Play consumers. However, congestion externalities dissipate approximately 40% of these gains.

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

This paper's main contribution is to show the economic importance of congestion externalities online. Previous literature highlighting the benefits of product variety coming from online markets and digitization, such as Brynjolfsson, Hu and Smith (2003), and Aguiar and Waldfogel (2018), did not capture this limiting factor. Quan and Williams (2018) also identify the limited benefits consumers receive from additional product variety, although there, the main friction comes from heterogeneity in consumer preferences, rather than congestion affecting consumer product discovery.

This paper also shows that changes in platform design can increase product variety on the

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

platform.[3] It confirms theoretical predictions about platform design and entry (Anderson and Renault (1999), Bar-Isaac, Caruana and Cuñat Martinez (2012), Goldmanis et al. (2010)), and descriptive evidence (Brynjolfsson, Hu and Simester (2011), Zentner, Smith and Kaya (2013)).

Product assortment, as measured by the number of products, is a key competitive outcome of concern for regulators in most online markets.[4] A key novel trade-off coming out of my results is that platform-led or regulator-led policy that encourages entry benefits consumers, but also intensifies congestion externalities. Personalization and recommendation technology has improved since the time period studied in this paper. Many online markets (e.g., Amazon, video streaming services) have better such technology today than the mobile app market in 2012-2014. The effects generated by a platform's category redesign would likely be smaller today and in the future. Nonetheless, my findings serve as a useful upper bound on congestion effects that could be generated by future platform-led or regulator-led policy aiming to increase entry online.

The paper proceeds as follows: Section I provides an overview of mobile app market and the re-categorization event. Section II describes the data and presents some summary statistics. Section III presents the reduced form evidence. Section IV discusses the specification and estimation of the structural model and the welfare decomposition. The final section concludes.

I. Background

A. *Consumer Browsing on the Google Play Store*

When a consumer opens the Google Play Store on their phone circa 2014, their first choice is between games or non-game apps. After this, the consumer chooses a category to look at more specific game or non-game product types. Once they choose a category, they have several lists of apps to choose from. Bestseller-lists display apps with the largest number of downloads over a certain period of time.[5] 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 I.B). 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.

[3] Previous empirical literature primarily focuses on the effects of platform design on price competition and match quality - e.g., Fradkin (2017) and Dinerstein et al. (2018).

[4] 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.

[5] The exact algorithm determining the position of an app in the top lists is unknown, but it is related to downloads (Rayburn 2013).

Consumers have a search bar at the top of the screen, and can use it to find apps. They also receive some personalized recommendations on the home page of the Google Play Store. However, surveys of app consumers during and after my sample period also show that browsing the category structure, as opposed to using a search engine or targeted recommendations, is the primary way in which consumers discover new products.[6] The search function in the Google Play Store was unreliable (Hill 2014). Google introduced app indexing in 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 (Menuier 2017). The "personalized" recommendations primarily featured already popular apps that consumers likely already downloaded.

B. March 2014 Game Re-Categorization Event

From 2009 to 2014, the Google Play Store had six game categories, compared to eighteen game categories in the Apple iOS store. Anecdotal reports suggested one reason for this discrepancy was Google's smaller initial app variety in 2009. 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. As described above, this was not the case by 2014.

On December 9, 2013, Google announced that it was expanding the number of game categories in the first quarter of 2014 (Kellex 2013). The new classification matched Apple's existing eighteen category structure. 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 store re-design until March 17, 2014. Industry observers and developers were reportedly surprised both by the initial announcement and by launch delays (Crider 2014).

Since the re-categorization tripled the number of game categories, it immediately reduced the number of apps in each category. 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.[7] An app is more likely to feature in the rotating "featured" app lists and be visible to consumers.[8]

[6] 58% of Android consumers discover new products through "General browsing in the app store", according to a Forrester survey from 2012. A Google/ Ipsos survey from 2014, and other surveys from 2011 and 2016 also have similar findings. See Online Appendix A.2 for more.

[7] Given the small number of categories, choosing a category should have negligible cost for consumers. However, if the number of categories is large, the categories themselves could compete for scarce consumer attention, and increasing the number of categories could also make it harder for consumers to find products, for the same reason that increasing the number of apps does. That is not the case in this particular market. See additional discussions in Sections III.B and IV.D.

[8] The titles of categories also became 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. I show evidence for this in Online Appendix C.2. On the supply side, developers knew that if they produce such a game, there is a clear place for it to be discovered. This is especially the case since the new categorization structure already existed on the iOS Store for years at that point, and developers frequently multi-homed.

These changes were consistent with Google's primary motivation for the re-design: improving the consumer search experience. An industry observer described that "searching for general apps on the Play Store is an exercise in frustration" under the old six game category system (Crider 2014). In the PR announcement for the 2014 change, Google stated that the new game categories "mak[e] it easier for players to find games they'll love" (Whitman 2014). A Google blog post discussing a subsequent re-categorization in 2016 clearly stated that new app categories "improve the overall search experience" and "mak[e] them more comprehensive and relevant to what users are looking for today" (Google 2016). 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.[9]

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

II. Data

A. *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 (AppMonsta 2012-2014).[10] This is the first paper to use this dataset.[11] The data contains all information that consumers observe on the Play 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 supplement this data with scraped historical app rankings from Flurry.com (Flurry 2012-2014). Using 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.[12]

[9]Formal tests for downloads in Figure C1 and Online Appendix C.4.1, and for entry 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.

[10]Weekly aggregation is only used to predict app downloads (see Online Appendix B.2).

[11]Liu, Nekipelov and Park (2014) use an app dataset from the same provider for 2011-2012.

[12]See additional discussion in Online Appendix B.2.

B. Data Management

App-Type Classification. — I classify game apps into *types* based on their categories in the last period of the sample.[13] An app belonging to the Music category in the last period after re-categorization is defined to be a "Music" type game.[14] 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, apps could also be present in categories whose names do not represent their type. For example, in 2012, 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, as fewer than 0.2% of apps change categories in my sample.[15]

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.[16] These can range to millions of downloads.[17] 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.[18] 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.[19]

[13] Apps can also be classified into a larger number of types based on alternative criteria, as in Kesler, Kummer and Schulte (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.

[14] 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 types. See Online Appendix B.1 for a description of this approach. Robustness checks show that dropping these apps does not change the baseline 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 *all* apps into types. Results are quantitatively similar across the two classification approaches.

[15] 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.

[16] These download measures do not include updates, and they also do not double count the downloads of different versions of the app.

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

[18] Intuitively, I observe a new app ranked 1st in a category with a lower bound of 50,000 downloads, 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.

[19] 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

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 Online Appendix C.4.3).

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 IV 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

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

of Gandhi, Lu and Shi (2014).

III. Reduced Form Evidence of Congestion

A. Main Results

In this section, I use the March 2014 re-categorization event to test whether app store congestion - driven by the number of apps - has an effect on consumer demand and usage. My main demand outcomes are aggregate and app-level downloads, since I do not have direct data on app usage.[20] I have two identification strategies to test for the presence of congestion externalities.

My first strategy uses a heterogeneous treatment effects difference-in-differences approach. Store re-design affected game apps, but not non-game apps. The two groups of apps were separated in the store, and the timing of the policy was a surprise (see Section I).[21] I focus on a subset of game app-types.[22] 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.[23]

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 as Action games, and three times as many Card games as Casino games. This means that 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. After re-categorization, Card games had more other apps in their category than Casino games. If congestion exists and the number of apps in a category matters for discovery, re-categorization should immediately 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 for app-type c (or app j) at time t:

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

where δ_t are time fixed effects, $\delta_{(j)c}$ are app-type (or individual app) 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/apps and zero for non-game types/apps. Small Type$_c$ is a dummy equal to one for Action and Casino apps/app-types, and zero otherwise. The

[20] 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.4.2. Results are qualitatively and quantitatively similar to those in the main text, suggesting that in this setting, downloads closely correlate with usage.

[21] Formal tests for parallel trends for downloads are in Online Appendix C.4.1.

[22] Average download effect estimates on all game categories are in Table C1 in the Online Appendix.

[23] Ten app-types did not have pre-existing categories: Adventure, Board, Education, Family, Music, Role Playing, Simulation, Strategy, Trivia, and Word games. These apps were also affected by the policy through a different channel - improved informativeness of the search process. I discuss it more in Online Appendix C.2. 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.

key parameter in this regression is τ^2, identifying the average heterogeneity in treatment effects between Arcade and Action, and Cards and Casino games. The absolute baseline throughout is the non-treated group (non-game apps).

I estimate the regression both at the app-type and the individual app level. The sample includes the four game app-types described above and all non-game app-types. The sample period is the four months, from January 2014 to April 2014, to limit confounding changes in product assortment.[24]

Estimates from this regression, at the app-type and app levels, are in Columns (1) and (2) in Table 3. 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 20-30%, but downloads increase for small sub-categories by 100-120%. 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 Card and Casino games is not changing much between just before and just after re-categorization. What changes is the number of irrelevant alternatives that compete for scarce visible space, which decreases more for Casino than for Card games and more for Action than Arcade games.

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 such as Racing games, whose title was not changed, saw reductions in the number of apps. This was likely because some pre- policy "Racing" apps were better described as Music or Family games, and fit better into one of the new category spaces. As a result, for any game app, the number of other apps in their category declined immediately after re-categorization.[25] As before, 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.

I use this to provide further 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.[26] Differencing absorbs most time-invariant app characteristics, and I also control for other app characteristics like average app rating and app price. The estimating equation is:

[24]I focus on four months rather than just February and March. Since the re-categorization happened roughly in the middle of March, effects identified using February and March are likely incomplete. Results using only February and March 2014 are in Table C3. They are qualitatively identical, but quantitatively smaller than in Table 3. Null estimates for non-existent ("placebo") events before and after actual re-categorization are in Table C4.

[25]See Figure C2 in the Online Appendix for direct evidence of this.

[26]I also estimate the same regression with the difference in paid app prices 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 congestion, rather than a change in competition intensity. See Online Appendix C.8 for additional discussion.

Table 3—: **Reduced form Congestion Evidence**

Outcome Variable:	ln(Tot. Type Downloads) (1)	ln(App Downloads) (2)	Post/Pre Δ ln(App Downloads) (3)
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
			(0.003)
Unit of Observation:	App-Type	App	App
Sample Period:	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
Year/Month FE	•	•	
App-Type FE	•		
App FE		•	
App Controls		•	•
Observations	112	4,770,936	142,254
R-squared	0.953	0.984	0.879

Note: The sample period in all columns is January 2014-April 2014. Data in Column (1) consists of monthly observations at the app-type level. Data in Columns (2) and (3) consists of monthly observations at the app level. Columns (1) and (2) data include all non-game apps and Arcade, Action, Card and Casino games. Column (3) data includes all game apps. Outcomes for Columns (1) and (2) are the natural logarithms of downloads at each aggregation level. The outcome for Column (3) 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 February 2014. Controls include year and month fixed effects and app-type or app fixed effects depending on the column. Column (3) does not include app fixed effects, since observations are already 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 × Small Type" is a dummy variable equal to 1 during and after March 2014 only for Action and Casino games. "Post/Pre 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.

$$\begin{aligned} ln(Downloads_{j,Post}) - ln(Downloads_{j,Pre}) = \alpha(ln(NApps)_{jc,Post}) - ln(NApps)_{jc^*,Pre}) \\ + \beta X_j + \epsilon_j \end{aligned} \tag{2}$$

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.[27] $NApps_{jc,Post}$ and $NApps_{jc^*,Pre}$ are similarly the number of apps in the category of app j before and after re-categorization.[28]

Estimates of this regression are in Column (3) of Table 3. They show a statistically and

[27] In Table C3 in the Online Appendix, I show that the results hold using only February and March data.

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

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.[29] 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.6 I find a similar elasticity using longer-run variation in the number of apps in a category.

B. Discussion

The estimates above suggest the existence of variety-driven congestion externalities in the mobile app market. 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 (Emslie 2017). It also reflects consumer complaints about the existence of too many apps in the market, and the difficulty of findings apps (Haselmayr 2013). The presence of congestion raises questions about the welfare gains consumers experience from additional product variety. When variety increases in a market, despite the *existence* of additional products, consumers only have access to a limited subset of these. I quantitatively evaluate the importance of congestion effects on consumer welfare in Section IV.

The evidence presented above comes from a single natural experiment in a particular market, which introduces concerns about the generality of the results. Would the effects be different in a market with better personalization, or better search-bar technology? Would the effects be different if there were more categories?

A similar platform re-design that triples the number of categories would likely produce different effects in a market with hundreds or thousands of categories, such as Amazon. In a market with many categories, more consumer time and effort would be spent browsing *between* categories, as compared to *within* categories. In an extreme example, a market with a category for each product is analogous to a market with no categories at all. In such markets, the benefits of introducing more categories would be lower, or even negative. This suggests there is likely an optimal number (or range) of categories to display in the mobile app market (conditional on the number of apps), but the variation in my data does not allow me to say what this number (or range) may be. I leave this question to future research.

More generally, in the sense of informing future policy, these results likely represent an upper bound of the magnitude of congestion externalities in online markets. Platforms such as Amazon and Netflix are continuously improving personalization and recommendation algorithms, so that consumers do not need to engage in costly category-based browsing. In a market with better recommendation technology, or a better search function than

[29]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.

in the Google Play Store in the 2012-2014 period, the effects of re-categorization, and variety-based congestion, likely would be smaller. That said, despite large investments in targeting and recommendation algorithms, they are far from perfect. For example, despite video streaming services such as Netflix being at the forefront of personalized recommendation technology, the 2019 Nielsen Total Audience Report finds that US consumers spend between 8 and 10 minutes in choosing what to watch. Only a minority of consumers, fewer than one third, watch algorithmically recommended titles (Nielsen 2019). This is likely because consumers have idiosyncratic product preferences, which are difficult for platforms to predict.[30] Category browsing is still an important method of consumer discovery, meaning that categorization and variety-based congestion still plays an important role today, and moving forward.

IV. Model Based Evidence

In this section, I set up and estimate a demand model that allows me to measure changes in consumer welfare in the market while accounting for congestion costs. I then evaluate the effects of congestion on consumer welfare from additional variety using the re-categorization as an illustrative example.[31]

My demand model conceptually follows the consideration-set framework (Goeree 2008,Moraga-González, Sándor and Wildenbeest 2015, Honka, Hortaçsu and Vitorino 2017). This framework models the consumer search process in a setting where consumers are not fully informed about the attributes of each product in the market. Consumers choose a subset of products to "consider," and then choose a product to consume out of that subset. Using this approach, it is possible to separate two types of product characteristics: (1) characteristics that affect consumer awareness of a product, and the probability a product appears in a consumer's consideration set (e.g., the difficulty of finding a product; (2) characteristics that affect consumer consumption utility and choice conditional on awareness. I set up and estimate a simple linearized version of such a model below.[32] This demand model accounts for fundamental sources of product discovery frictions in the Google Play Store, and in 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.

[30]It is possible to make an analogy to personalized pricing / first degree price discrimination. Although in principle fully personalized pricing is possible, it is not attempted by firms because of noisy consumer preferences (Dubé and Misra 2022).

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

[32]A non-linear version of a consideration-set model following Moraga-González, Sándor and Wildenbeest (2015) is in Appendix D.2.

A. Demand Model

Consumer i chooses to download app $j \in \{1, 2, .., N\}$ in market/period t. The utility she receives from downloading an app j of type c is:[33]

$$\begin{aligned} u_{ijct} &= \delta_{jct} + \gamma ln(N_{c^*(j)t}) + R_{jct}\kappa + \zeta_{ict} + (1-\sigma)\epsilon_{ijct} \\ &= X_{jct}\beta + \xi_{jct} + \gamma ln(N_{c^*(j)t}) + R_{jct}\kappa + \zeta_{ict} + (1-\sigma)\epsilon_{ijct} \end{aligned} \tag{3}$$

where X_{jct} and ξ_{jct} are unobservable and unobservable app/app-type characteristics that affect consumption utility, respectively. Observable app characteristics include its average rating in period t, whether it's paid or free, app price for paid apps, and app "age" (months on the market). Characteristics that do not enter into δ_{jct} capture utility changes coming from changes in app discovery costs, rather than consumption utility. This distinction is conceptual in the main text, but in the model in Online Appendix D.2, these variables enter the model non-linearly and are distinct from consumption utility.

$ln(N_{c^*(j)t})$ is the number of apps in the *category* of app j. As discussed in Section III, 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 appears in a space that is visible to consumers.[34] c^* represents the actual observed categorization of apps in the store rather than app-type (c), although after re-categorization c and c^* coincide for app j.

R_{jct} are additional shifters that proxy discovery costs, capturing competition for unobservable store space and scarce consumer attention. One such shifter is an app's downloads in the previous period - capturing inter-temporal externalities in the mobile app market, and reflecting the design of many online markets. An app with more downloads in period $t-1$ is more likely to appear on a best-seller list and remain highly visible in period t. R_{jct} also includes time-varying app-type specific fixed effects. Each app-type has two dummies, one before and one after re-categorization.[35] 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.

ϵ_{ijct} is a consumer/app/market specific demand shock with an iid EVT1 distribution (mean zero, standard deviation normalized to 1).[36] ζ_{ict} is a consumer/app-type/market

[33] An app's "type" is defined according to the classification described in Section II.B.

[34] This is similar to the modelling in Ackerberg and Rysman (2005). $N_{c^*(j)t}$ 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 III, I show that the number of apps in a category matters for downloads when the number of substitutes is held constant. In Online Appendix C.8, I also show that changes in the number of apps in a category do *not* affect paid app prices. If the number of apps in a category primarily changes competition intensity, it should affect paid app prices.

[35] Formally, $\sum_c \Theta_{c,Pre}(D_c \times D_{Pre}) + \sum_c \Theta_{c,Post}(D_c \times D_{Post})$, 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. Θs are coefficients on the combinations of dummies.

[36] Consumer preference for variety comes primarily through this shock. This is a reasonable assumption for a

specific demand shock, such that $[\zeta_{ict} + (1-\sigma)\epsilon_{ijct}]$ 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.[37] The consumer can also pick an outside option of not downloading anything and receive $u_{i0} = \epsilon_{i0}$.

Based on the assumed distribution of idiosyncratic demand shocks, the market share of app j, belonging to app-type c in market t is defined as:

$$s_{jct} = \frac{exp(\frac{R_{jct}\kappa+\gamma ln(N_{c^*(j)t})+\delta_{jct}}{1-\sigma})}{\sum_{j'\in c} exp(\frac{R_{j'ct}\kappa+\gamma ln(N_{c^*(j')t})+\delta_{j'ct}}{1-\sigma})} \frac{\left[\sum_{j'\in c} exp(\frac{R_{j'ct}\kappa+\gamma ln(N_{c^*(j')t})+\delta_{j'ct}}{1-\sigma})\right]^{1-\sigma}}{\sum_{c'\in\{1,..,C\}}\left[\sum_{j''\in c'} exp(\frac{R_{j''c't}\kappa+\gamma ln(N_{c^*(j'')t})+\delta_{j''c't}}{1-\sigma})\right]^{1-\sigma}} \tag{4}$$

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

B. Demand Estimation and Results

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

$$ln(\frac{s_{jct}}{s_{0t}}) = X_{jct}\beta + \gamma ln(N_{c^*(j)t}) + R_{jct}\kappa + \sigma ln(s_{j|c,t}) + \xi_{jct} \tag{5}$$

where s_{0t} is the market share of the outside option.[38] $s_{j|c,t}$ is the within-app-type market share in period t. I estimate the model by solving this equation for the structural error term ξ_{jct} and interacting it with instruments to form the GMM criterion function. As discussed above, X_{jct} includes app-age fixed effects, R_{jct} includes time-varying app-type fixed effects, and I also include year and month fixed effects and developer fixed effects to control for additional unobservable heterogeneity.

The key parameters I look to identify in this model are: the price coefficient for paid apps (one of the variables in X_{jct}), β_{price}, γ and σ. β_{price} is important for measuring the welfare effects of changes to product variety or characteristics in dollar terms. σ identifies consumer substitution patterns across apps, and γ identifies the congestion externality.

The three variables associated with the coefficient of interest suffer from endogeneity. App prices and within-type market shares are likely correlated with unobservable product

market with many minimally horizontally differentiated product versions such as "Angry Birds Space" and "Angry Birds Star Wars."

[37]There could be additional unobservable heterogeneity in preferences for quality. Estimating a random coefficients model using aggregate app data with hundreds of thousands of products and a small number of markets is computationally challenging. There is also likely not enough variation in market shares to identify the distribution of unobservable heterogeneity (Berry, Linton and Pakes 2004, Armstrong 2016).

[38]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.

quality (ξ). Products with higher ξ have more demand, higher prices and higher within-type market shares, meaning that simple OLS estimates of the price coefficient and σ should be biased. Similarly, there may be correlation between the number of apps in a category and some unobservable category/time varying demand shocks, biasing OLS estimates of γ. I address these concerns using instrumental variables.

I use characteristics-based instruments to address the endogeneity of app prices and within-group market shares: average ratings, app size (in MB), and number of screenshots of other type c apps, excluding app j. These characteristics proxy for competitors' average quality. For example, the quality of competing apps is excluded from the utility consumers receive from downloading a paid app j, but it does affect app j's price through competition.

The instrument for $N_{c^*(j)t}$ has to capture supply-side entry cost shocks and exclude demand-side factors. The number of apps in a given category is correlated over time - apps rarely exit, and those who entered in previous periods are persistently present in the market. However, demand shocks could also be correlated over time - consumers may persistently like certain games or game categories in both periods t and $t+1$. To create my instrument, I estimate 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.[39] I then use the residual of that regression as the IV. Intuitively, realized period t total downloads and 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 should be driven by supply side cost shocks.[40]

I estimate the model using data on all game apps from March 2012 to December 2014 (excluding March 2014). Results for the main parameter of interest are in Table 4.[41] There are four columns in the table. Each column has a different set of instrumental variables. Column (1) estimates an OLS version of the model with no instruments. Column (2) includes characteristics-based IVs for prices and the within-group market share, but no IVs for the number of apps in a category. Column (3) is the opposite: it includes the instrument for the number of apps in a category, but no IVs for prices and within-group shares. Column (4) includes all IVs.[42] Standard errors are clustered at the app level throughout.

Comparing estimates across columns suggests that the instruments generate appropriate variation to target the relevant endogenous variables. In Columns (1) and (3), σ is very close to 1, which is unintuitive and generally suggests the presence of attenuation bias. The price coefficient is similarly very small, implying that consumers are price-inelastic and that the vast majority of paid-app marginal costs are negative. This is also problematic, and together, these estimates suggest endogeneity is a substantial concern.

[39] Estimates of this supporting regression are in Table D2 in Online Appendix D.

[40] This instrument is conceptually similar to the cross-sectional Hausman-Nevo IV specification of Dubois and Lasio (2018).

[41] Additional estimates of other parameters that are less important for the counterfactuals in this paper are presented in Online Appendix Table D1.

[42] In addition to the instruments described in the main text for the main variables of interest, I also instrument for lagged app downloads (a variable in R_{jct}) using functions of earlier lags in app downloads (e.g., q_{jt-2}, q_{jt-3}).

Table 4—: **Demand Model Parameter Estimates**

	(1)	(2)	(3)	(4)
Parameter Estimates				
γ	0.053	-0.066	-0.432	-0.393
	(0.001)	(0.008)	(0.003)	(0.026)
σ	0.963	0.742	0.989	0.709
	(0.000)	(0.013)	(0.000)	(0.027)
β_{price}	-0.001	-0.355	-0.001	-0.836
	(0.000)	(0.051)	(0.000)	(0.111)
Instruments				
Characteristics IVs		•		•
N Apps IV			•	•
Fixed Effects				
Paid App Dummy	•	•	•	•
App Age	•	•	•	•
App Rating	•	•	•	•
Year/Month	•	•	•	•
Developer	•	•	•	•
App Type × Pre/Post Re-Categorization	•	•	•	•
Other Controls				
Lagged Downloads	•	•	•	•
Other App Quality Proxies	•	•	•	•
Observations	4,584,281	4,190,276	4,152,328	4,152,147

Note: 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." Year/Month FE include separate dummmies for each year (2013 and 2014, relative to a 2012 baseline), and each month (Feb/Mar..., relative to a Jan baseline) in the sample period. Lagged downloads are the downloads of app j in period $t-1$. Other app quality proxies include the number of screenshots for app j in period t, the size of the app in MB, and a dummy for whether the app has a video preview. Instruments in Columns (2) and (4) include the average ratings of other apps in the same app-type, the average number of screenshots of other apps in the same app-type, and the average size of other apps in the same app-type. Instruments in Columns (3) and (4) include the residual of a regression of the current number of apps in a category on the lagged number of apps in a category (see Online Appendix Table D2 for estimates). Additional IVs in all four columns, for lagged app downloads, include functions of further lags in app j downloads (2 and 3 periods before period t). Standard errors are clustered at the app level.

β_{price} and σ estimates in Columns (2) and (4) suggest that characteristics based-IVs appropriately resolve these concerns. In Column (4), the price coefficient implies a median paid game demand price elasticity of over 2, which is realistically elastic. Estimates of σ in Columns (2) and (4) are 30% smaller than in Columns (1) and (3).

Similarly, the Column (1) estimate for the γ coefficient is positive, which is unintuitive given the findings in Section III, and is likely caused by endogeneity. The estimate in Column (2) is negative, but not nearly as negative as in Columns (3) and (4), where the appropriate instrument is incorporated into the estimation procedure. Results in Columns (3) and (4) are consistent with reduced form evidence and suggests that an increase in the number of apps in app j's category reduces demand and consumer utility. Put otherwise, there are congestion externalities on consumer demand in the mobile app market.

In the welfare analysis below, I use estimates from Column (4), but I also tested several alternative specifications of the model for robustness. Results are in Table D3 in Online Appendix D. For example, I estimate a simple nested logit model with no discovery frictions. This model does not fit the data well, producing σ estimates above 1. I also allow new apps to have different discovery frictions than incumbent apps. There are no statistically significant differences in key parameter estimates between the two types of apps. Last, I estimate a non-linear GMM model based on the search and demand model developed in Appendix D.2. Results are qualitatively similar to those in the main text with respect to the main parameters of interest.

C. Welfare Costs of Congestion

In this section, I first show that re-categorization increases entry rates for mobile games and product variety in the market. Then, using the estimated model from the previous section, I calculate the welfare gains consumers experience from the new products and the costs of congestion.

RE-CATEGORIZATION AND ENTRY. — I first test for whether re-categorization affected app entry. In each month, I observe the number of unique entrants of each app-type.[43] Similar to Section III, I estimate the following regression for app type c at time t:

$$ln(\text{N Entrants}_{ct}) = \tau(Game_c \times Post_t) + \delta_c + \delta_t + \epsilon_{ct} \tag{6}$$

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 6 is τ, which captures the treatment effect of re-categorization on the outcome. Entry is a long-run decision, so I estimate the regression in Equation 6 using the entire 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.[44]

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

[43] I only consider new apps that appear in 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.

[44] 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.

Table 5—: **Entry Difference in Differences Estimates: Average Effects**

Outcome Variable:	ln(N Entrants) (1)	ln(N Type Entrants) (2)
Games × Post	0.342	0.556
	(0.068)	(0.095)
Games	-6.914	
	(1.893)	
Unit of Observation	Game/Non-Game	App-Type
Time Period	Jan 12 / Dec 14	Jan 12 / Dec 14
Sample	All	All
Year/Month FE	•	•
App-Type FE		•
Observations	70	1,470
R-squared	0.997	0.975

Note: The sample period in all columns covers January 2012 to December 2014. Sample in Column (1) includes monthly observations at the Game/Non-Game level. Sample in Column (2) 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. Controls include year and month fixed effects, game/non-game fixed effects for Column (1), and app-type fixed effects for Column (2). 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 Column (1) and clustered at the app-type level in Column (2).

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.[45]

In Online Appendix Figure C3, I show that the effects begin two months after the re-categorization announcement in December 2013. This is consistent with the average timing of app development, and suggests that new apps enter in response to the re-categorization. Figure C3 also shows that entry changes peak the month after re-categorization, but are present and persistent throughout the post-policy period. Online Appendix C.7.1 shows the mechanism through which the entry effects operate. Entry is primarily driven by the new categories in the store, which improved consumers' information and reduced discovery costs. Entry is also bigger for app-types which saw greater immediate reductions in congestion.[46]

Over all, the estimated entry changes reflect theoretical predictions from search models, such as Anderson and Renault (1999) and Chen and Zhang (2017). As product discovery costs fall, consumers are more likely to search more, and producers respond by entering new products to benefit from increased visibility. Results are also consistent with previous descriptive comparisons between high and low discovery cost markets. Brynjolfsson, Hu and Smith (2003) shows that online retailers have between 10 and 30 times more products

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

[46] As for the downloads results in Section III, these estimates are particular to the small baseline number of app categories. As discussed in Section III.B, in a market with many categories, such as Amazon, more consumer time and effort would be spent browsing *between* categories, as compared to *within* categories. There, the effect of introducing more categories on entry would likely be smaller.

than brick-and-mortar retailers. Aguiar and Waldfogel (2018) notes that the number of new music products between 2000 and 2008 tripled. However, neither of these papers separates changes in search costs from other changes in technology, such as production costs.

The increase in entry should satisfy consumer preferences for variety, but it should also increase congestion externalities. Based on demand parameter estimates in Section IV.B, this' should restrict the benefits consumers receive from the new products. I quantify these welfare effects below.

Welfare Decomposition. — Based on the demand model in Section IV.A, changes in consumer surplus are the difference between expected utilities under different sets of products. I convert these changes into dollar values using the price coefficient from Table 4. A consumer who downloads app j of type c receives baseline utility $\hat{\delta}_j = X_j\hat{\beta}$. That consumer also receives/pays congestion costs based on the number of other apps in the category $(\hat{\gamma}ln(N_{c^*(j)t}))$ and other "discovery costs" $R_{jct}\hat{\kappa}$. The last two terms are linearly additive to $\hat{\delta}_j$. Based on the demand model from Section IV.A, expected consumer utility is:

$$EU = ln\left(1 + \sum_{c\in\{1,..,C\}}\left[\sum_{j\in\Omega_c} exp\left(\frac{\hat{\delta}_j + R_{jc}\hat{\kappa} + \hat{\gamma}ln(N_{c^*(j)})}{1-\sigma}\right)\right]^{1-\sigma}\right) \tag{7}$$

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

I calculate three relevant expected utility measures. First, I calculate expected utility using *the actual set of apps available in December 2014*, the last period available in the sample. I refer to this expected utility as EU_1. Second, I calculate expected utility using a set of apps in a *counterfactual December 2014 market without re-categorization*, EU_2. I calculate this counterfactual choice set by randomly removing apps that entered between March 2014 and December 2014 from the December 2014 choice set.[47] Reduced form estimates from Section III tell me how many entrants to remove. I remove entrants to match the aggregate entry treatment effects from Column (1) of Table 5.[48] I repeat this exercise 500 times to ensure randomization is not driving results. It is important to note that for a given product j, there is no difference in $\hat{\delta}_j$ or in R_{jc} between the factual and counterfactual markets. Only the set (and number) of products in the market changes.

[47] An alternative approach to doing this 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 assumption. 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. Estimating an entry model with many players also introduces severe computational challenges.

[48] 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.

The difference $\frac{EU_1 - EU_2}{-\beta_{price}}$ is the overall effect of additional entry on consumer welfare. This is a "net" effect. Re-categorization changes Ω, the set of products in the market, and also $N_{c^*(j)}$, the congestion measure. To separate the two, I calculate EU_3, which is based on Equation 7 where Ω comes from the factual December 2014 market, and $N_{c^*(j)}$ comes from the counterfactual December 2014 market. Put otherwise, it measures expected consumer utility under the factual choice set, but with counterfactual congestion based on the counterfactual number of entrants. $\frac{EU_3 - EU_2}{-\beta_{price}}$ isolates the *gross* entry effect of only changing the choice set while holding congestion constant. $\frac{EU_1 - EU_3}{-\beta_{price}}$ isolates the negative welfare effect due to the increase in congestion between the counterfactual and factual markets.

Table 6 shows the decomposition of consumer surplus changes. I show mean entry and congestion effects based on the randomized procedure outlined above.[49]

Table 6—: **Estimated Changes to Monthly Per-Consumer Surplus ($)**

1) Mean Gross Entry Effect	$EU_3 - EU_2$	0.057
2) Mean Entry Congestion Effect	$EU_1 - EU_3$	-0.024
3) Net Entry Effect	$EU_1 - EU_2$	0.033

Note: This table shows estimates of average per-consumer welfare effects using demand model estimates from Column (4) of Table 4. Numbers shown are averages from 500 simulations. Full distributions of simulation effects are in Figure D3 in the Online Appendix.

Without accounting for congestion externalities, the additional entry spurred by the redesign of the store has a large effect on consumer welfare. Row (1) of Table 6 shows that *gross* consumer welfare increases by 5.7 cents per consumer per month, adding up to 68.4 cents of additional per-consumer welfare in a year. This is equivalent to approximately one third of the price of the average paid game app. Across all 100 million Google Play Store consumers in 2014, gross annual welfare gains add up to nearly $70 million.

However, row (2) of Table 6 shows that the congestion externalities associated with the increase in product variety reduce consumer welfare gains. Consumer lose approximately 40% of gross variety welfare gains because of increasing variety-driven congestion costs. In absolute terms, losses from additional congestion are 2.4 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 $29 million per-year for all Google Play Store users.

Over all, consumer welfare still increases. Row (3) shows that the net increase in consumer surplus is 3.3 cents per-consumer per-month, or approximately 40 cents per-year (and

[49]Figure D3 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.

$40 million across all consumers), relative to a counterfactual world where re-categorization does not happen. Nonetheless, these estimates show that congestion externality costs in online markets are economically substantial, and consumers do not fully benefit from additional product variety in the market.

D. Model and Counterfactuals Discussion

The demand model and counterfactuals presented above (and in Online Appendix D) abstract from several elements of the consumer search process in the Google Play app store. It does not explicitly model the costs consumers incur while browsing *between* categories, only the costs of browsing *within* categories. In a market with many categories, consumers may spend a substantial amount of time choosing which category to look through. I lack consumer level data to be able to estimate such costs, but it is reasonable to assume that in the mobile app market in 2012-2014, they are likely very small compared to the costs of browsing within categories. They also do not change for the relevant counterfactual welfare decomposition in Section IV.C. Some estimates from the model speak to their magnitude/ importance: the model measures average changes in consumer utility from downloading a type c app before and after the re-design using time-varying app-type fixed effects. Within app-type differences in these fixed effects capture many changes, including better consumer information about category content, but they should also include changes in category browsing time/ costs. If the costs of browsing between categories become large and important after the re-design, the difference in fixed effects should be negative. However, estimates of these differences in Online Appendix Figure D1 are uniformly positive. This suggests that category browsing costs are either small in magnitude, or that they do not increase substantially. In that sense, abstracting from these in the model should not affect the main results and conclusions for this particular setting.

The model also abstracts from platform personalization on the Google Play Store. In reality, all consumers see personalized app recommendations on the Google Play Store home page. Some consumers may immediately see a targeted product that caters to their preferences, and purchase it immediately without engaging in a costly browsing process. Other consumers may find what they are looking for quickly using the search bar function. For those users, variety-driven congestion does not matter. However, based on consumer surveys (see Online Appendix A.3), and on statements by industry experts (see Section I), this was likely a minority of app consumers during (and even after) my sample period. Abstracting from this element should not substantially affect my estimates or conclusions for this market.

These issues also raise broader questions about the extent to which the findings here generalize. There are two possible generalizations to be made from the findings of this paper. First, my results suggest that past findings of the value of added product variety online (e.g., Brynjolfsson, Hu and Smith 2003 and Aguiar and Waldfogel 2018) may be overstated, because they do not account for congestion externalities. Papers in this literature primarily use pre- 2012 data from online markets, where targeting and personalization technology was plausibly worse than in the Google Play Store. The magnitude of my estimated congestion externalities, therefore, is appropriate, and could even be a lower bound.

Second, my findings have implications for future platform- led or regulator- led policies that affect online entry. These could involve platform re-design and re-categorization, as in my illustrative case, or other policies such as direct subsidies to entrants.[50] A policy increasing entry would increase consumer welfare through new variety, but the benefits would be mitigated by increased congestion. Future policies are likely to occur on platforms with substantially better search and personalization technology than the Google Play app store in 2012-2014. As such, the counterfactual results suggesting that congestion costs dissipate 40% of welfare gains from additional variety are likely an upper bound. At the same time, personalization technology is unlikely to ever be perfect, as consumers have idiosyncratic product preferences that are difficult to predict.[51] As discussed in Section III.B, recent surveys show that streaming video consumers still spend substantial time browsing platforms, and that most consumers do not choose the algorithmically provided option. Since streaming video platforms such as Netflix are at the forefront of personalization and recommendation technology, this implies that variety-driven congestion should still have an effect today and in the future.

V. Conclusion

This paper examines the role of variety-based congestion in online markets. Product assortment is the key competitive outcome of concern for regulators of online markets. Understanding consumer welfare changes in response to changes in assortment is important. I use new data from the Google Play mobile app store, taking advantage of a re-design of the game section of the store. Reduced form estimates using changes in downloads show that when the re-design increased the number of categories, reducing the number of apps-per-category, downloads increased. Downloads also increased by more for apps that were in less populous categories. This points to the existence of variety-based congestion externalities.

I set up a structural demand model that accounts for variety-driven congestion externalities. This allows me to decompose consumer welfare changes from variety into "gross" changes in utility due to the larger assortment, and welfare changes driven by congestion. Increased mobile game entry following the Google Play Store re-design serves as an illustrative example. Entry increased by 34% after the re-design, because of improvements in consumer discovery technology. Relative to a counterfactual market where the re-design did not happen, each consumer received an additional 40 cents per-year from the new product assortment. This adds up to nearly $40 million additional aggregate welfare per year. However, absent congestion, welfare increases would have been even higher. Increased congestion takes away 40% of gross variety gains. These findings suggest that previous es-

[50] Alibaba, for example, provided direct subsidies and free services to firms operating on its platform (Xinhua 2014).

[51] As discussed previously, this is analogous to the inability of firms to first-degree price discriminate online (Dubé and Misra 2022).

timates in the literature were overstating consumer benefits from additional online product variety.

Policy-makers are concerned with platforms foreclosing potential entrants and reducing variety, for example, by making product discovery difficult for consumers (Mullins, Winkler and Kendall 2015). My findings show that welfare gains from variety are counteracted by increasing congestion externalities. In my application, welfare changes coming from congestion are not enough to overwhelm the variety effect, and consumers benefit. However, welfare changes from increased congestion are still economically significant and annually add up to tens of millions of dollars of lost consumer surplus. Personalization and recommendation technology has improved since my sample period, and it is substantially better in many online markets today (e.g., Amazon, video streaming services). Nonetheless, these findings serve as a useful upper bound on the magnitude of congestion effects that could be generated by future platform-led or regulator-led policy aiming to increase entry online. Empirical analysis of platform incentives to invest in better search and personalization technology - or change their design to facilitate consumer discovery - in the presence of congestion externalities, should be a fruitful direction for future research.

REFERENCES

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

Aguiar, Luis, and Joel 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.

Anderson, Simon P, and Régis Renault. 1999. "Pricing, product diversity, and search costs: a Bertrand-Chamberlin-Diamond model." *RAND Journal of Economics*, 30(4): 719–735.

AppMonsta. 2012-2014. "Details for all apps in the Google Play Store." *API access through "https://api.appmonsta.com/v1/stores/android/...". Accessed in 2015.*

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

Bar-Isaac, Heski, Guillermo Caruana, and Vicente Cuñat Martinez. 2012. "Search, design and market structure." *American Economic Review*, 102(2): 1140–1160.

Berry, Steve, Oliver B Linton, and Ariel Pakes. 2004. "Limit theorems for estimating the parameters of differentiated product demand systems." *The Review of Economic Studies*, 71(3): 613–654.

Bruemmer, Joseph. 2012. "Why consumers can't find you online." *iMediaConnection.*

Brynjolfsson, Erik, Yu Hu, and Duncan 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, Erik, Yu Hu, and Michael 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, Yongmin, and Tianle Zhang. 2017. "Entry and welfare in search markets." *The Economic Journal*, 128(608): 55–80.

Crider, Michael. 2014. "Google Play Store Game Category Expansion Is A No-Show For February, Support Page Now Says March." *Android Police.*

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan. 2018. "Consumer price search and platform design in internet commerce." *American Economic Review*, 108(7): 1820–59.

Dubé, Jean-Pierre, and Sanjog Misra. 2022. "Personalized pricing and customer welfare." *Journal of Political Economy.*

Dubois, Pierre, and Laura Lasio. 2018. "Identifying industry margins with price constraints: Structural estimation on pharmaceuticals." *American Economic Review*, 108(12): 3685–3724.

Emslie, Rowan. 2017. "The Complete Guide To App Store Categories." *PrioriData.*

Flurry. 2012-2014. "Rankings of apps in the Google Play Store." *Downloaded through "flurry.com/...". Accessed in 2015.*

Fradkin, Andrey. 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).*

Gamet, Jeff. 2018. "Apple Hampers Mac App Store Discoverability by Removing Several Categories." *Mac Observer.*

Gandhi, Amit, Zhentong Lu, and Xiaoxia Shi. 2014. "Demand Estimation with Scanner Data: Revisiting the Loss-Leader Hypothesis." mimeo.

Goeree, Michelle Sovinsky. 2008. "Limited information and advertising in the US personal computer industry." *Econometrica*, 76(5): 1017–1074.

Goldmanis, Maris, Ali Hortaçsu, Chad Syverson, and Önsel Emre. 2010. "E-Commerce and the Market Structure of Retail Industries*." *The Economic Journal*, 120(545): 651–682.

Google. 2016. "Introducing new app categories – From Art to Autos to Dating – to help users better find your apps." *android-developers.googleblog.com.*

Haselmayr, Melanie. 2013. "App Discovery: Why Can't Anyone Figure This Out Yet?" *Forbes AllBusiness.*

Hill, Simon. 2014. "Google needs to tackle its app discovery problem." *Android Authority.*

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino. 2017. "Advertising, consumer awareness, and choice: Evidence from the us banking industry." *The RAND Journal of Economics*, 48(3): 611–646.

Kellex. 2013. "All Sorts of New Game Categories Coming to Google Play in February 2014." *DroidLife.*

Kesler, Reinhold, Michael Kummer, and Patrick Schulte. 2020. "Competition and privacy in online markets: Evidence from the mobile app industry." Vol. 2020, 20978, Academy of Management Briarcliff Manor, NY 10510.

Liebowitz, Stan J, and Alejandro Zentner. 2020. "The Challenges of Using Ranks to Estimate Sales." *Available at SSRN 3543827.*

Liu, Yongdong, Denis Nekipelov, and Minjung Park. 2014. "Timely versus quality innovation: The case of mobile applications on iTunes and Google Play." *NBER Working Paper.*

Menuier, Bryson. 2017. "A Case Study: Is App Indexing For Google Worth The Effort?" *Smashing Magazine.*

Moraga-González, José L, Zsolt Sándor, and Matthijs R Wildenbeest. 2015. "Consumer search and prices in the automobile market."

Mull, Amanda. 2019. "There Is Too Much Stuff." *The Atlantic.*

Mullins, Brody, Rolfe Winkler, and Brent Kendall. 2015. "Inside the U.S. Antitrust Probe of Google." *The Wall Street Journal.*

Nielsen. 2019. "The Nielsen Total Audience Report: September 2019."

Padla, Reah. 2015. "Google Play now shows "Family" Category." *Android Community.*

Quan, Thomas W, and Kevin R Williams. 2018. "Product variety, across-market demand heterogeneity, and the value of online retail." *The RAND Journal of Economics*, 49(4): 877–913.

Rayburn, Scott. 2013. "Google discloses how search for Google Play works for the first time." *AdWeek.*

Whitman, Ryan. 2014. "[GDC 2014] Google Announces New Game Developer Tools And Game Gifting (But Not The Kind You Were Hoping For)." *Android Police.*

Xinhua. 2014. "Alibaba provides subsidies for exporters." *China Daily.*

Zentner, Alejandro, Michael Smith, and Cuneyd Kaya. 2013. "How video rental patterns change as consumers move online." *Management Science*, 59(11): 2622–2634.