Money and Privacy - Android Market Evidence*

Michael Kummer†
Georgia Institute of Technology &
Centre for European
Economic Research (ZEW)

Patrick Schulte‡
Centre for European
Economic Research (ZEW)

February 29, 2016

Please do not Quote without Permissions from the Authors

Abstract

We shed light on a money-for-privacy trade-off in the market for smartphone apps. Developers offer their apps cheaper in return for greater access to personal information and consumers choose between lower prices and more privacy. We provide evidence for this pattern using data on 300,000 mobile applications which were obtained from the Android Market in 2012 and 2014. We augmented them with information collected from Alexa.com and Amazon Mechanical Turk. Our findings show that both supply and demand take apps' ability to collect private information, measured by apps' use of privacy-intrusive permissions, into account: (1) Cheaper apps use more privacy-intrusive permissions; (2) Installation numbers are lower for apps which ask for intrusive permissions. (3) Circumstantial factors, such as the reputation of app developers, mitigate the strength of this relationship. Our results survive several robustness checks, including the use of panel data analysis, the use of selected matched "twin"-pairs of apps and applying three independent classifications and various alternative measures of privacy-intrusiveness.

JEL Classification: D12, D22, L15, L86

Keywords: Privacy; Mobile Applications; Android; Permissions; Supply and Demand of Private Information.

*We are grateful to Irene Bertschek, Jörg Claussen, Anindya Ghose, Avi Goldfarb, Shane Greenstein, Sang-Pil Han, Andres Hervas-Drane, Fernando Luco, Bertin Martens, Markus Mobius, Martin Peitz, Arnold Picot, Imke Reimers, Rahul Telang, Bernd Theilen, Catherine Tucker, Hal Varian, Frank Verboven, Joel Waldfogel, Michael Ward, Simon Wilkie, Manfred Wittenstein, Pinar Yildirim, Pai-Ling Yin, Michael Zhang and Christine Zulehner for valuable comments and helpful advice. We thank the participants of the 12th ZEW ICT conference 2014, EARIE 2014, WISE 2014, IIOC 2015, 6th SEARLE Internet Search and Innovation 2015, SEEK Digitial Economy Workshop Torino 2015, NBER Summer Institute 2015, and Sevilla Apps Economy Workshop 2015. We thank Niklas Duerr, Florian Hofbauer and Steffen Viete for their extremely useful research assistance.

†221, Bobby Dodd Way, #222 Atlanta, GA, 30308, U.S.A Email: michael.kummer@econ.gatech.edu.

‡P.O. Box 103443, D–68034 Mannheim. Email: schulte@zew.de.

1 Introduction

In this paper, we use data on 300,000 mobile apps from Android's Market Place to show to which extent private information has taken the role of a second currency in the market for mobile applications. We highlight that developers trade greater access to personal information for lower prices and consumers choose between lower prices and more privacy. Understanding this money-for-privacy trade-off is useful, since the recent rise of smartphones and tablet PCs has been the most significant change in the market of end user computational devices over the last decade.[1] To a large degree, this is due to the emergence of a market for mobile applications (henceforth "apps") which allows users to tailor their devices to their personal needs. However, while apps have become a very important market, little is known about how this two-sided market actually works. Especially, little is known about the role of privacy in the context of apps, despite the enormous potential of smartphones to collect various, unseen sorts of private information about consumers. This potential raises multiple important questions: How do suppliers trade direct revenues for more usage and the possibility of getting access to private data? Do users avoid "data greedy" apps consciously?

We exploit Google's unique policy of informing users about an app's potential access to private information and combine it with data on 300,000 mobile apps to answer these questions. Our inference is based on a full cross-section, a monthly panel, a 2-year panel and a set of carefully selected and matched pairs of apps. Our data were repeatedly collected from the Android Market in 2012 (over 6 months) and in 2014. The information covers all publicly available app-specific information including e.g. information on apps' number of installations and their price. Most importantly, we exploit the fact that Android's Operating System enforces a unique information policy with regard to how an app can access personal user information. Specifically, users are confronted with the complete list of access rights that an app will have and must acknowledge that they are granting these permissions *before* the installation. We discern 136 permissions that apps could request in 2012 and recorded the permission requirements of each app. Finally, we augmented our data with information from Alexa.com and with data we collected on Amazon Mechanical Turk to add background information about app providers and permissions.

Our results document a money-for-privacy trade-off on both sides of the market for mobile applications: (1) App developers clearly ask for more and for more privacy-intrusive permissions if they offer an app for free than if they offer it for a (higher) price (2) We can find permissions which are not necessary for the functionality of an app. These point to a conscious use of the apps' ability to collect private information for monetization purposes; and (3) consumers take this trade-off into account and reduce their demand for apps

[1] According to OECD's estimates, the number of users who access the internet via mobile devices is currently surpassing the number of users using a fixed line. See e.g. OECD Broadband Database (2012).

which ask for very privacy-intrusive permissions. Our analysis is based on a panel data set, a "long panel" with long term outcomes (2 years later) and a set of carefully selected and matched pairs of apps. To control for unobserved heterogeneity and show that our results are robust to using panel data analysis. Our results consistently emerge across these different subsets of data and in various specifications. We can also show that the patterns are consistent for both the demand and the supply side. The findings persist for different ways to quantifying intrusiveness (Dummy, number of permissions), and they are also not rejected in a long panel and a matched dataset of app-pairs. The results are also robust to using three alternative measures of intrusiveness: a classification by Sarma et al. (2012), one by Google and one based on classifiers we hired on Amazon Mechanical Turk. Also, our findings are robust towards using the number of apps, instead of a zero-one classifier, both in a parameterized and a non-parametric specification.

In addition to these robustness checks, we analyze "sister-pairs" of apps: two versions of the same app, where one is offered for free and the other one for pay. We manually identified pairs where the paid version does not offer any additional service, and collected data on almost 600 such pairs. These pairs are used to analyze the supply side behavior of app developers and the price they charge for additional privacy. Note also that the paid version can serve as a technological reference to identify redundant permissions. Any permission X that is present in the free, but *not* in the (potentially more powerful) paid version, will not be related to *functionality*, but rather to monetization. The paid twin version is the proof of concept that the same app can run without the critical permission.

Finally, we scrutinize which factors modify the money for privacy trade-off. We augment our data with additional information from outside sources, such as Alexa.com which allows us to study the role of mitigating factors. We analyze which circumstantial factors influence the strength of our findings: We analyze whether consumers react differently (1) to permissions labeled as malicious by Google, (2) to privacy-intrusive permissions common in an apps' category, or (3) if an app is offered for a positive price. We also analyze how the relationship of interest differs (4) if the developer of the app has a good reputation, (5) if a consumer is relatively young and (6) if the app belongs to more sensitive categories (such as medicine or business apps).

Our findings highlight a privacy for money trade-off in the market for mobile applications, which appears to be considerably weaker for well known brands, than for new or unknown apps. This unfavorable discrepancy might constitute a significant barrier to entry of new developers which could result in a less than optimal innovative performance of the market for mobile apps. If the result is a consequence of consumers' lack of knowledge regarding privacy sensitive permissions, policy makers might want to consider user-friendly certification procedures by outside parties (e.g. a traffic light scheme) to help to reduce this entry barrier.

The paper follows the usual structure. In section 2 we discuss the related literature, and in section 3

we provide information on how we obtained the data. In section 4 we discuss our estimation strategies and section 5 features our results and robustness checks. We discuss the implications of our findings and the limitations of our work in section 6, before concluding the paper in section 7. The appendix contains tables and figures and an online appendix discusses additional details about the data, and presents further results or robustness checks.

2 Relationship to the Literature

At the latest since the revelations of Edward Snowden the role and value of private data has become a central theme of the (economic) public debate.[2] Hence, a small but growing stream of research has begun to provide urgently needed empirical evidence on the role of private information for supply and demand in online markets. Preibusch and Bonneau (2013) analyze data collection policies of internet sites and find that the intensity with which they collect data varies substantially. Relatedly, a series of recent studies analyzes how privacy policies affect users of social networks or the success of targeted advertisement (Goldfarb and Tucker (2011), Tucker (2012), Johnson (2013a), Tucker (2014), Aziz and Telang (2015)). They show that restrictions on the usage of private data in advertisement substantially reduced targeting effectiveness, which resulted in lower revenues for the content site, but also highlighted that privacy policies have an important effect on consumer behavior.[3]

On the demand site, e.g. Gross and Acquisti (2005) study demand for privacy in social networks. Acquisti et al. (2013), using a field experiment, examine consumer preferences for privacy and highlight the sensitivity of privacy valuations to contextual factors. Goldfarb and Tucker (2012) use information from an online survey to study how privacy concerns have changed over time and find evidence that they have increased, especially for older people. More recently e.g. Marthews and Tucker (2014) study the effect of government surveillance on internet search behavior. Their results suggest that internet users, after Snowden, reduce demand for search terms which might get them in trouble with the US government.

Regarding the role of private information in app markets, existing research is based on experimental and survey data. Grossklags and Acquisti (2007) and Tsai et al. (2011), contrasting consumers' willingness to pay to protect privacy and their willingness to accept for giving away their personal information, show that the willingness to pay is much lower. Also a recent survey based study (Savage and Waldman (2013)) found that consumers' self-reported willingness to accept giving away the personal information, that is typically shared with developers, is near 4 USD. The choice architecture of the platform affects smartphone users'

[2] For a survey of theoretical and empirical scientific studies concerned with the economics of privacy, see Acquisti et al. (2015).

[3] In contrast to empirical evidence, several theoretical models have analyzed the role of private information in online markets. In such models the knowledge about personal preferences of an agent can be used to price discriminate (Wathieu (2002), Taylor et al. (2010)). Also, firms can use customer information, such as the purchase history, to charge personalized prices in settings of electronic retailing (Taylor (2004), Acquisti and Varian (2005), Conitzer et al. (2012)). An alternative way to use the personal information is the context of direct marketing, which may result in costly efforts to avoid ads (Johnson (2013b), Hann et al. (2004)). Increasing the cost of anonymity can benefit consumers, but only up to a point, after which the effect is reversed (Taylor et al. (2010)). Taken together, these models see reduced privacy as a source of inefficiency. However, as shown by Spiegel (2013), e.g. in the context of software production, providing free software in a bundle with (targeted) ads could be welfare improving if the cost of producing software is relatively low. Also, of course, due to reduced privacy many valuable services can be provided "for free" and can that way create benefits for users. This ambiguity is e.g. studied recently by Casadesus-Masanell and Hervas-Drane (2015) who analyze a situation where suppliers compete in privacy. In their model, they analyze the effect of privacy sensitive information in a setting of competition in two-sided markets, as it was pioneered by Armstrong (2006) or Rochet and Tirole (2003).

willingness to pay premiums to limit their personal information exposure (Egelman et al. (2013)).[4] Recent theoretical work looked at privacy as a second currency by analyzing situations where suppliers compete in privacy Casadesus-Masanell and Hervas-Drane (2015). We build on the theoretical and survey based evidence and extend it by analyzing large scale data based on observed market transactions.

With respect to the functioning of app markets some evidence exists. Yin et al. (2014) and Davis et al. (2014) study innovation in app markets. Carare (2012) is concerned with the impact of bestseller ranks on app demand. The impact of large scale promotions on the sales and ratings of mobile apps is examined by Askalidis (2015). Most closely related to our work is Ghose and Han (2014) who estimate the demand for selected (top-rated) apps. They focus on 300 top-rated apps on the Android and Apple platforms and compare them. Our study differs in three important ways from their previous work. First, we add a new focus by analyzing the role of privacy in such markets, for both, demand and supply. Second, we can increase the scope of the analysis because we observe the complete set of apps that was available in the Android Market in summer 2012 ($N = 300,000$). Third, we observe a discrete measure of real downloads, rather than approximating demand via the sales ranks.

We contribute to the literature by studying the money-for-privacy trade-off in a recently emerged and increasingly important online market as well as by providing evidence on the functioning of app markets based on large scale market-transaction data. We observe detailed information on the permissions (rights) that the app requests before installation and can use them to shed light on apps' "demand" for personal information as well as to provide evidence about how users' installations are related to the these permissions. Finally, we analyze the behavior of suppliers who offer the same app (i) for pay, but with limited access to personal information and (ii) for free, but with greater access to the user's personal data.

[4]Studies having a more technical focus investigate the precise meaning of certain permissions and what they imply for the privacy of the device's owner (see e.g. Chia et al. (2012),Sarma et al. (2012)). Other studies investigated how dangerous apps can potentially be (e.g. Chia et al. (2012) or Fahl et al. (2012)).

3 The App Market

In 2007, Apple Inc. introduced the highly successful iPhone. It can be considered the starting point of a radical transformation replacing the use of mobile phones by smartphones. The availability of large app ecosystems can be considered one of the main competitive advantages of the iPhone and its successors. Apps allow users to tailor their devices to their needs, allowing for a multitude of uses next to the traditional use of a phone. Inspired by the success of Apple, 2008 then saw the release of the first phone using Google's Android Operating System. Although Android's adoption was relatively slow at first, it started to gain widespread popularity in 2010, and now dominates the market (in most countries). According to IDC (2015) already in 2012 the Android OS reached a market share of around 75 percent. Nowadays (in 2015) the revenue of the mobile app store is around 40 billion US dollar and is expected to reach 100 billion US dollar in 2020 (Annie (2016)).

In addition to the Operating System, Google also introduced its own platform for the distribution of apps. The platform was originally named "*Android Market*," but was renamed "*Google Play Store*" mid 2012. Since then it does not only serve as a distribution channel for apps, but in addition also for Books, Music and Movies and newspapers. In 2012 it had around 400,000 apps available, a number which increased to around 1.5 million in 2015. Google provides a categorization of thirty categories which can be sorted into overarching meta categories, such as Education, Entertainment, Games, Tools&Personalization, Lifestyle, Health and Business. Table 3 shows how we classified the categories into seven meta categories. In 2012, the largest meta category was Tools&Personalization (69,372 apps), which contained weather and transportation apps (surprisingly not games). The smallest category were Health and Business related apps (8,255 and 11,686 apps respectively).

The central feature of the Android app ecosystem for this paper is its permission system. This system is specific to the Android Market and provides the setting in which the money-for-privacy trade-off can be meaningfully studied. First, developers can choose among predefined standardized blocks of information (henceforth "permissions"), which include the access to information about users' location, their communication, their browsing behavior etc. They know that their app must declare which personal user information it can access and must request the necessary permissions. Second, to install an app, users have to accept all requested permissions *before* installation.[5] Specifically, a list of permissions' names was provided alongside a short explanation of each permission. Users had to accept this list and to explicitly acknowledge that they are granting these permissions to proceed with the installation. Alternatively they could cancel the installation if they felt uncomfortably about the set of permissions requested. Note that explicit consumer consent to the

[5]The most recent version of the Android OS (Android 6.0 Marshmallow) is an exception (see below).

set of permissions is different from Apple's system, where this information is not made explicit. In its essence this procedure remained stable since 2012, and is still in place today despite the fast growth of the Android Market.

In 2012 developers could choose among 136 predefined permissions.[6] This illustrates how much and how many diverse types of information app developers can potentially collect about app users. Since then smaller modifications with respect to the way these permissions are displayed to the user have been introduced over time. Before 2014, the list of permissions provided permission names next to short explanations of the permissions. In 2014 a more aggregated form of illustration was introduced which only displays names of permission groups (but allows opening a more detailed dialogue containing more information on a respective permission group). Still, before proceeding with the installation process, the this permission list has to be approved. Very recently (for the most recent version of the Android OS (Android 6.0 Marshmallow), Google introduced a major change in its permission system which now allows users to withdraw individual permissions from an app after the installation.[7]

Finally, there are four important channels how developers can monetize their apps. According to AppBrain (2016), around 20 percent of the apps are paid apps, whereas the remaining apps are for free upon installation.[8] Alternative revenue channels are in-app advertisment, in-app purchases and data trade. The importance of these alternative revenue channels has been relatively stable since 2012 except for in-app purchases, which were introduced shortly before our period of observation. In 2012, when we collected our data, in-app purchases were hardly used. By 2015 however, the freemium model based on in-app purchases has become prevalent. In this model, which virtually did not exist in 2012, the apps can be installed for free, but important functions must be unlocked for a fee.[9] The two other channels, in-app advertisement and data trade were already commonly used. While in-app advertisement is deemed more acceptable by some users, data-trading is clearly the more privacy-sensitive way of creating revenue from an app's use.

[6]This number has remained more or less constant and equals nowadays 137 permissions (see `http://developer.android.com/reference/android/Manifest.permission.html`).

[7]The resulting effects cannot be evaluated in this paper since only a small share of users has access this version of the OS.

[8]Developers receive 70 percent of the app price, and 30 percent go to distribution partners and operating fees (see `https://support.google.com/googleplay/android-developer/answer/112622?hl=en`).

[9]Only in 2011, Google added in-app billing to Android Market, allowing apps to sell in-app products (see `http://android-developers.blogspot.de/2011/04/new-carrier-billing-options-on-android.html`).

4 Data and Descriptive Evidence

In the first subsection we describe how we collected data from Google's Android Market - one of the two largest platforms for mobile applications. The second subsection provides descriptive statistics and three stylized facts about the money-for-privacy trade-off in the market for mobile apps, as they appear in the raw data.

4.1 Data

We extracted all publicly available information on most apps available in 2012 and 2014. We collected the data monthly from April to October 2012 and once in 2014. The repeated data collection in 2012 allows us to use panel data methodology. The additional wave from 2014 was gathered to analyze long-term outcomes, such as installation growth over 2 years. Our data set covers nearly the full population of products available in 2012 (around 300,000 apps). Figure 4 in the Appendix shows the design of Google's Android Market in 2012 which corresponds to the information we were able to collect. To study our research questions, we need mainly three types of information: a demand measure, a price measure and a measure of apps's ability to collect private information. In the following section we introduce each of these measures as well as the core control variables.

Main Outcome Variables: Our main demand measure is the number of installations of an app. Our data set contains direct information on the total number of installations (i.e. sales) for each app. Our demand measure is provided in discrete form (17 levels, e.g. 1-5 installations, 6-10 installations, 11-50 installations, etc.). This is an improvement over most previous internet-based data sets, where demand variables have to be approximated via sales ranks (see e.g. Chevalier and Mayzlin (2006), Garg and Telang (2013) or Ghose and Han (2014)). In addition we use the number of ratings of an app, which is available as a continuous measure, to improve upon this demand measure. Specifically we exploited the discrete download measure to predict continuous downloads based on the number of ratings and we use this proxy in our panel analysis. Equally important to a demand measure are information on prices, for which we have precise information (in Euro) for each app.

Identifying Privacy-Intrusiveness of Apps: To measure apps' ability to collect private information, we take advantage of the fact that Google's Android Market provides precise insights into the permissions an app uses. This unique feature allows us to understand in a detailed way which rights an app has, and thus which functions it can perform, including functions which allow an app to collect private information about the

app user. In 2012, Google had defined 136 different of such permissions an app could use, such as e.g. 'read SMS or MMS', 'fine (GPS) location', 'read browser data', etc. More examples are provided in Figure 5 which illustrates the way the permissions were displayed and described in the Android Market. These permissions have to be declared in the app description and have to be accepted by the app user before installing the app.[10]

Among those 136 permissions, some are innocuous with respect to the privacy of the user, while others grant an app access to sensitive information. To identify those permissions being a risk to the privacy of the user, we use four alternative permission classifications. Our main classification (1) is derived from previous research by Sarma et al. (2012). The three alternative classifications are (2) a category-specific modification thereof, (3) Google, and (4) hiring 400 classifiers at Amazon Mechanical Turk.

Our workhorse definition of privacy-intrusive permissions follows Sarma et al. (2012), who analyze the benefits and risks of Android app permissions and classify them according to different risk types. We follow them in classifying 26 permissions as critical among which 13 are considered as being a risk to privacy. [11] Based on this classification, we construct our main variable of interest ($D_{Privacy}$), which is a dummy equal to one if an app uses at least one of the 13 privacy-intrusive permissions and zero otherwise. To capture the intensity of an apps's ability to collect private information, in addition, we make use of the number of privacy-intrusive permissions per app.

Three alternative classifications of privacy-intrusive permissions: We tested robustness by employing three alternative definitions of privacy-intrusiveness. The first one modifies our baseline definition by classifying only those permissions within an app category as privacy-intrusive, which are both (a) defined as privacy-intrusive by Sarma et al. (2012) and (b) which are used below-average within an app category. The idea is that permissions which are used rarely within an app category are atypical for this kind of app such that the respective permission might be less likely required for the apps' functionality but for collecting information about users. To give an example, the permission 'read contact data' is very common in business apps, social apps, sports apps or productivity apps, but may be less necessary for a weather app, a medical app or an app for personal finances.

The second alternative classification uses Google's own classification of 'potentially malicious permissions'. For 38 of the 136 available permissions, Google's official permission description includes a note that the respective permission might be 'potentially malicious,', i.e. it might harm the user of this app. To identify a privacy-intrusive subgroup of those 38 potentially harmful permissions, we combine this classification with

[10] We use the standardized short explanation to inform users about the permission's meaning by Google.

[11] For the permission *read calender* we were not able to collect information, such that we only have information on 12 privacy-intrusive permissions.

that of Sarma et al. (2012) and form two groups: potentially-malicious privacy-intrusive permissions and not-potentially-malicious privacy-intrusive permissions. The former group is defined as the group of permission which are classified as potentially malicious by Google and as privacy-intrusive by Sarma et al. (2012).

For a third and independent robustness check we use a categorization that we obtained by hiring 450 workers on Amazon Mechanical Turk. To classify the permissions they were presented with a randomly selected subset of permissions. For each permission the workers were asked how likely they would hesitate to proceed with the installation of an app, if noticing that it was requiring them to grant the permission. While we do not consider the self-reported likelihood as a very reliable absolute measure we can use the relative measure. If a permission was generally likely to incite hesitation we classified it as very problematic, while if it was relatively unlikely to raise concerns, we classified the permission as unproblematic.

Table 2 summarizes all classifications applied and describes each privacy-intrusive permission. In addition, it shows how we grouped the privacy-intrusive permissions into four subgroups: location-, profile-, communication- and ID-specific permissions. Even more details about the data collection are provided in the online appendix.

Control Variables: Next to our main variables, we also observe a rich set of app-specific characteristics relevant for explaining app supply and demand: the app category, the number and average of ratings, code size, android version, developer-specific information (name of developer, number of other apps, top deveoper status, etc.), the app's description (length, number of screenshots, video). Also, we can use the section "users who viewed this also viewed," which provides us with information on related apps. We identify app-specific competitors and construct three additional control variables: (i) the average price of competitor apps, (ii) competitors' average installations, and (iii) the average rating of competitor apps.

Categorizing apps by type: Finally we used Google's categorization to identify seven overarching categories of apps. Table 3 shows how we classified Android's thirty categories into seven overarching meta categories. This was merely done for simplified representation for category-specific results.[12] The table shows all the 30 categories in Android's Playstore sorted by their size. Moreover it indicates to which of our seven overarching categories the categories were added. We defined Education, Entertainment, Games, Tools&Personalization, Lifestyle, Health and Business. Surprisingly, in 2012, Games were not the largest category of apps. Instead, the largest resulting meta category is Tools&Personalization (69,372 apps), which also contain weather and transportation apps. The smallest are Health and Business related apps (8,255 and

[12]Category-specific privacy sensitive permissions are computed based on the 30 underlying categories. The category-specific results do not depend on this classification, the coefficients, for all 30 categories estimated separately are available upon request.

11,686 apps respectively).

4.2 Descriptive Evidence

In this section we present summary statistics for the most relevant variables and provide descriptive evidence on three stylized facts that we uncover in our paper.

4.2.1 Summary Statistics - Three Data Sets

Table 1 provides an overview over the most important variables and describes our three most important data sets. The descriptive statistics in the table are shown in groups of two columns, were the left column shows averages for free apps and the second column for paid apps.

Table 1: Summary statistics of three datasets

	Cross Section		Pairs		Panel	
	Free	Paid	Free	Paid	Free	Paid
Outcome Variables						
Installations (in 1000)	104.62	1.83	294.04	2.90	292.65	12.05
Pred. Num. Installs.	82.45	2.05	152.41	2.22	213.19	7.82
Average Rating	3.90	3.95	3.86	4.11	4.03	4.14
Permissions						
$\#_{TotalPerm.}$	4.49	2.10	4.26	1.91	7.18	5.07
$\#_{CriticalPerm.}$	3.06	1.33	3.21	0.97	4.42	2.85
$\#_{PrivacyPerm.}$	1.14	0.37	0.94	0.33	1.85	0.95
$D_{Privacy}$	0.50	0.22	0.52	0.22	0.75	0.46
$D_{PrivCatSpec}$	0.25	0.07	0.21	0.03	0.39	0.14
$D_{MaliciousPrivacy}$	0.33	0.11	0.22	0.09	0.49	0.29
$D_{NonmaliciousPrivacy}$	0.45	0.18	0.49	0.18	0.69	0.38
$D_{MTurkSP}$	0.51	0.23	0.53	0.23	0.76	0.49
$D_{MTurkVP}$	0.19	0.07	0.12	0.10	0.27	0.20
$D_{MTurkEP}$	0.05	0.01	0.02	0.02	0.05	0.05
$D_{Internet}$	0.82	0.44	1.00	0.28	0.95	0.72
D_{Ads}	0.58	0.21	0.86	0.08	0.80	0.46
D_{Other}	0.41	0.28	0.35	0.25	0.64	0.58
App Characteristics						
Price	0.00	2.17	0.00	1.18	0.00	3.25
App Version	20.60	10.43	26.41	21.54	23.88	11.80
Size (in KB)	2345.07	4265.88	2245.35	1925.27	3440.90	7189.29
Length Description	716.89	957.97	965.54	855.17	1023.50	1608.83
Number Screenshots	3.35	3.48	3.90	3.87	4.19	5.27
Dummy: Video	0.10	0.09	0.13	0.12	0.12	0.27
Dummy: Top-Developer	0.01	0.01	0.01	0.01	0.02	0.04
Apps by Developer	123.64	215.77	13.08	13.08	53.81	33.38
Average Installations of Developer	99.35	43.96	76.92	142.42	193.72	206.33
Observations	233811		992		61927	

Notes: The table provides an overview over the most important variables, and shows the corresponding descriptive statistics for the three main datasets in this paper. For each dataset we show two columns, where the left column shows averages for free apps and the second column for paid apps. The first two columns show the permission usage in the entire cross section. The second pair of columns (Col. 3-4) show the descriptive statistics for apps that are available as a free and paid twin of the same app. Colums 5 and 6 show the panel dataset of apps that change permissions over time. Developer specific variables were computed excluding the current observation. The mean of the predicated number of installations for the cross-section and for the pairs are computed using a slightly reduced sample where observations with zero number of ratings were dropped.

The first two columns show the permission usage in the entire cross section of apps containing of 233,813 observations.[13] Free apps are installed more often, and they have a lower average rating. Crucially, both the number of privacy critical permissions used in an average app and the likelihood that apps use such permissions is higher among the free apps. For example, free apps use on average 3.06 privacy relevant permissions, whereas paid apps use only 1.33 such permissions on average. Similarly 25% of the free apps have at least one privacy relevant permission, that is not "category specific" (not usually used in this category of apps), while only 7% of the paid apps are found to use such permissions. We obtain the same result when comparing the presence of permissions that Google flags as having "potentially malicious use," and when applying the classification that we obtained from hiring 450 classifiers on Amazon Mechanical Turk. A third of the free apps uses "potentially malicious" permissions (only 11% of the paid), and 19% of free apps use a very problematic permission, as classified by our hired classifiers, while only 7% of paid apps do so.

The second pair of columns (Col. 3-4) show the descriptive statistics for our most rigorously matched pairs of apps. These apps are available as a free and paid twin of the same app, by the same developer.[14] App-pairs are interesting, because introducing sensitive permissions in the free but not in the paid version is the most explicit instance of a "money for privacy" trade-off.[15] Pairs are also quite useful to the researcher, because we can find sensitive permission that are certainly redundant: The paid version is the proof that the same app can run without the critical permission. Note that this redundancy continues to be valid even if the paid version of the app offers *more* features than the free version, if only the paid app does not offer fewer!

For the pairs shown in columns 3-4 of Table 1 we manually verified two criteria. First their description and code length had to be exactly the same (or longer for the free app), and second, they should only differ in permissions and price. These apps are downloaded more frequently than the average app in the cross section, which indicates that they are generally more successful. The ratings are very similar and expose a higher spread and the price is lower than the average paid app. For the usage of permissions we find very similar patterns as in the full cross section, with maybe slightly less permissions in the payable "twins." Especially for the classifications we obtained ourselves the gap becomes more narrow and even vanishes for "extremely problematic" permissions.

Lastly (Colums 5-6) we show our panel dataset of apps that change permissions over time. Using panel

[13]The discrepancy between those 233,813 apps and the full app population of around 300,000 apps is mainly due to two reasons: for around 20,000 apps we were technically not able to collect information about them, whereas the remaining 50,000 apps are dropped from our sample since they were not available in some of the subsequent monthly waves we used for our panel analysis. To ensure a balanced panel, we decided to restrict both the panel data and the cross-section data to apps which were observable throughout the full observation period.

[14]These pairs were identified using a word processing algorithm which identifies app pairs having the same name except for one of the following addings: 'free', 'paid', 'lite', 'full', 'demo', 'pro', 'premium', 'donate', 'trial', 'plus'.

[15]A typical, though not ideal, pair are a "lite" or "demo" version (free) and the full version (for pay) of the same app, where the free version collects more data.

data will allow us to deal with unobserved app-specific characteristics. Especially unobserved quality or functionality are important in our context, since certain features naturally require permissions from the phone. If a running app has the feature of challenging your running friends, by sending them your most recent workout, it will need to access your contacts for that. At the same time, this feature will correlate with how attractive the app is in general, which will drive up downloads. Using panel analysis we can explore the relationship between downloads and permissions, when we control for unobserved app-specific characteristics. The apps in our panel dataset are more frequently downloaded and also have slightly higher ratings. These apps also use permissions more heavily than the average app from the general dataset. The difference between free and paid apps is less pronounced in relative terms but remains the same in absolute terms.

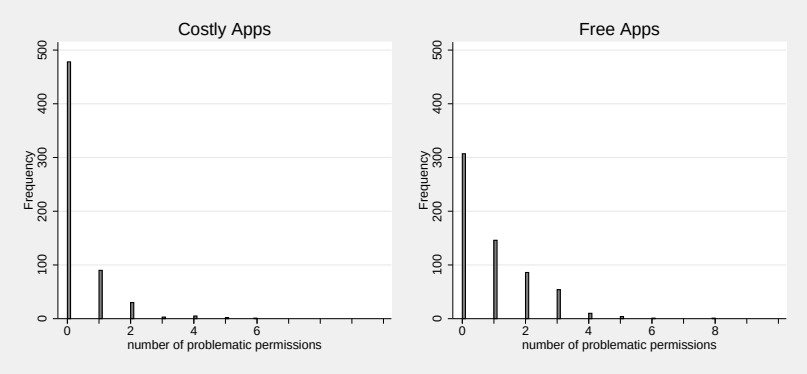
Already the simple analysis of the summary statistics of these three data sets suffices to reveal a consistent pattern: No matter how we look at the data, free apps always use both more permissions, and more 'privacy relevant' permissions than paid apps. This pattern is found even when matching the pairs of apps that come from the same developer and have the same functionality. Beyond the three data sets in Table 1, we also gathered an additional full wave of data in 2014. This allows us to put together a long term panel of two waves, and we also use that data to analyze survival rates and 2-year growth of the downloads based on the 2012/2014 differences (shown in Table 10).For reasons of space the dataset is not described here, but is available upon request.

4.2.2 Three Stylized Facts

In the following we provide evidence on three stylized facts. First, we present unambiguous evidence that some apps indeed use permissions which are not necessary for their functionality. We conjecture that these redundant permissions are used for monetization purposes. Second, we highlight that free apps are much more likely to use privacy-intrusive permissions than apps which are for pay. This indicates a negative relationship between price and permission use. Third, the number of installations is negatively related to privacy-intrusive permissions, indicating that demand is negatively related to such permissions. The following three paragraphs elaborate on each of these stylized findings.

Redundant Permissions: We use the sample of selected app pairs to study whether apps indeed use permissions without need for the functionality of an app. Iin this data set the paid version can serve as a technological reference, because the paid version can safely be assumed to provide *more* (or at least similar) functionality than the free one. Hence, any permission that is present in the free version but not in the paid sister is redundant *for functionality*, and can be expected to be related to monetization. Figure 1 contrasts the distribution of privacy-intrusive permissions in the free and paid versions of apps. The left side of the

Figure 1: The distributions of privacy sensitive permissions (free vs. paid services).



Notes: The figure contrasts the distribution of privacy sensitive permissions (according to Sarma et al. (2012)) in the free and paid versions of apps. The selection of pairs equals those, which are likely to differ only in that the free version uses ads. The left side of the graph shows paid apps, whereas the right one shows the distribution for free applications. Free apps are more likely to use privacy sensitive permissions.

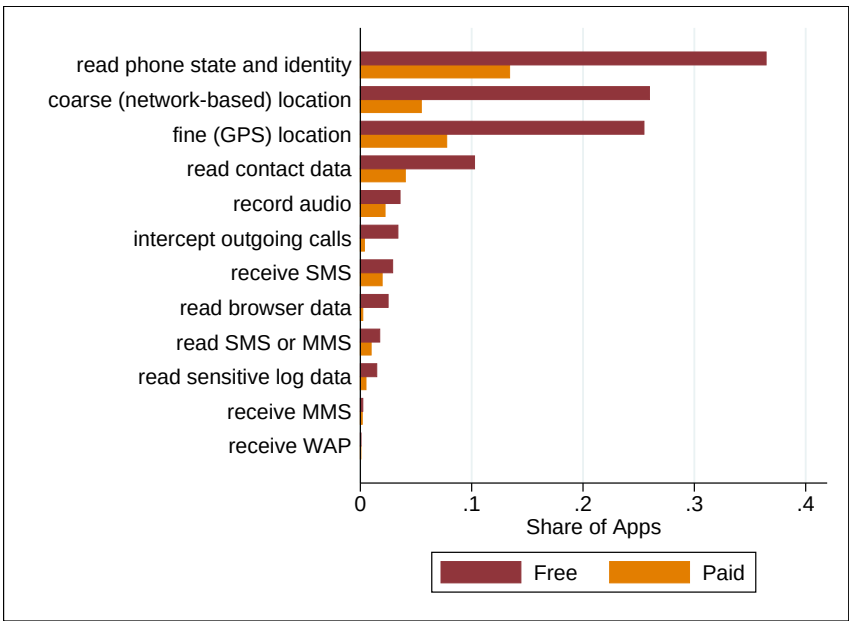
graph shows paid apps, whereas the right one shows the distribution for free apps. As can be seen, the share of "clean" twin-apps without any problematic permissions is much higher for paid apps. Meanwhile, many free apps use one or more problematic permissions, which, considering the special nature of app pairs, are not necessary for functionality. We conclude that several of the free versions turn out to be using redundant and privacy sensitive permissions.

Price and Permissions: Next we show that developers charge lower prices when requesting more permissions. Figure 2 shows the most frequently used privacy sensitive permissions, and contrasts the share of free and paid apps that use them. The chart highlights that these permissions are considerably more common in free apps. We conclude that developers indeed choose between offering an app either for a low price (or for free) but with many privacy-intrusive permissions or offering an app for a higher price but with a lower number of privacy-intrusive permissions.

Demand and Permissions: Figure 3 illustrates our third stylized fact: Apps using privacy-relevant permissions seem to be installed less often than apps which do not use such permissions. The figure displays coefficient estimates of twelve privacy-relevant permissions. These are obtained by running simple, descriptive regressions for each of the permissions. The dependent variable is the log-normalized number of installations and the only explanatory variables are the respective permission and, as a control, the total number of permissions an app uses. The coefficients are negative and significant for most of the privacy-intrusive permissions. Apps which use such permissions are installed less often than apps which do not use them.

In the subsequent sections we provide rigorous evidence, especially on the latter two relationships. Above

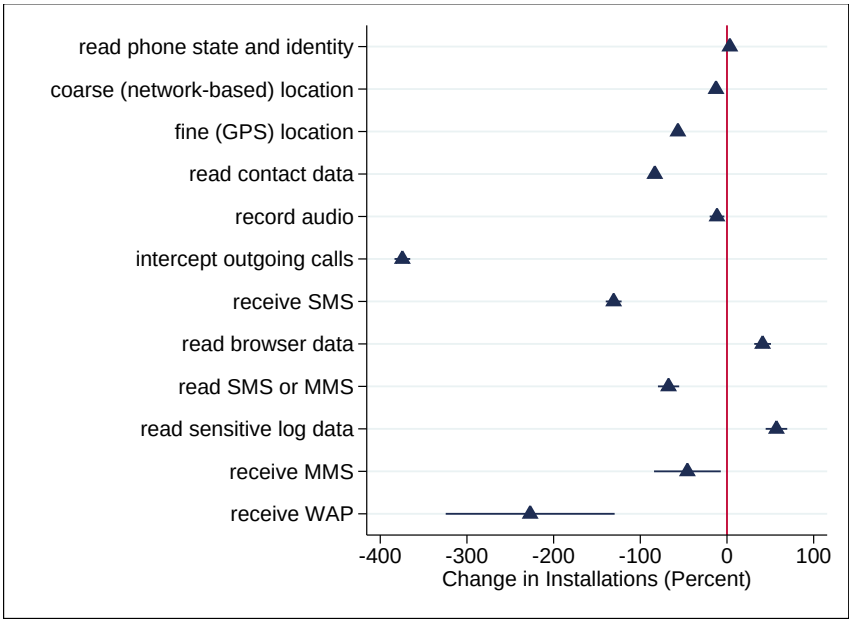
Figure 2: Frequency of Critical Permissions - Free & Paid



Notes: This Diagram shows how ofthen privacy-relevant permissions are used, both by free and paid apps. Free apps use such permissions much more frequently than paid apps.

all, we will focus at establishing that these stylized facts hold if we control for potential omitted factors.

Figure 3: Number of Installations and Permissisions



Notes: This Figure displays the coefficient estimates of the privacy-relevant permissions. They are obtained with a simple estimation where the dependent variable is the log number of installations and the explanatory variables are the total number of permissions and the privacy-relevant permission under study. The Figure suggests that for most of the privacy-relevant permissions, apps which use such permissions are installed less often than apps which do not use such permissions.

5 Estimation

This section discusses estimation. The first subsection presents an overview over our estimations of the supply side. Subsequently, we discuss estimation of the demand side, using the cross-section and our alternative approaches to tackle potential unobserved heterogeneity.

5.1 Supply Side Estimations

To provide insights on the role of private information for the supply side. We run two types of correlational regressions for app developers. First, we examine the choice of the business model. Specifically we analyze whether a developer offers an app for free or for paid relates to its use of privacy-relevant permissions. Our estimation equation equals:

$$D_i^{Price} = \alpha + \beta D_i + \theta X_i + \varepsilon_i, \tag{1}$$

where D_i^{Price} is a dummy equal to zero if an app is for free and equal to one if an app is for paid. The dummy indicates whether an app uses specific permissions, whereas X_i are control variables. A negative coefficient for β indicates that an app which used the specific permission D_i is β less likely to be for paid. Alternatively, it indicates that free apps are more likely to come with a privacy-sensitive permission.

In a second specification we restrict our sample to paid apps and study how the price level relates to the use of specific permissions. The estimation equation then equals:

$$\ln Price_i = \alpha + \beta D_i + \theta X_i + \varepsilon_i, \tag{2}$$

where the dependent variable equals the log price of an app. Here a negative coefficient β indicates that for paid apps, the price level is lower if the specific permission is used.

Panel Data Analysis: Finally, we attempt to control for unobserved heterogenity by using a panel specification. We put together a monthly panel covering our observation period of five months. Analyzing the within variation focuses on changes in price or the business model within the same app and allows us to control for unobserved heterogeneity. However, this approach is limited by the small number of apps who change price or switch business model in our period of observation. For the business model we estimate the following fixed-effects model:

$$D_{it}^{Price} = \alpha_i + \beta D_{it} + \theta X_{it} + \varepsilon_{it} \tag{3}$$

We estimate the analogous relationship for the price of apps (provided they request payment).

5.2 Demand Side

We first present results of our baseline specification, where we analyze the relationship between privacy-intrusiveness and app demand. Subsequently, we present our approaches to tackle potential unobserved heterogeneity.We use panel estimation and a dataset of matched "sister-pairs" of apps, where the same app is provided for pay and for free. Furthermore we consider alternative proxies of demand.

Baseline Specification: To analyze the relationship between app demand and permissions which allow an app to collect personal information we use econometric analysis. We apply a straightforward empirical demand specification, which models demand for an app as a function of its permissions, its price and other observable characteristics. Our main specification is based on the cross-section sample. The models were estimated using both, simple OLS and 2SLS that accounts for the endogeneity of the price. We estimate the following baseline model:

$$Demand_i = \alpha + \beta D_i + \gamma P_i + \theta X_i + \varepsilon_i \tag{4}$$

$Demand_i$ for app i is approximated by the number of installations of an app. Demand is a function of a vector of permission group dummies D_i, the app's price P_i and a set of observable characteristics X_i, including the log of the average rating, the app version, the app category, the length of the app description, the number of screenshots, a dummy for the existence of a video, a top-developer dummy, the logarithmic number of apps of its developer, the average number of installations of the app's developer's apps, the minimum and the maximum compatible android version as well as information about the app's competitors' characteristics. ε_i is the error term.

Endogeneity of Prices: In what follows we shall focus on the simple OLS, because the estimated parameters of interest do not differ very much after instrumenting. However, we tested whether our coefficients of interest (the role of privacy sensitive permissions) are affected by the high likelihood of endogeneity in the price variable. We use 2SLS specifications, and instrument the endogeneous price variable:

$$Demand_i = \alpha + \beta D_i + \gamma P_i^{IV} + \theta X_i + \varepsilon_i \tag{5}$$

The instruments used are cost shifters (size of the code) and the price of the closest substitutes (apps by competitors). Note also that for free apps, the price is constant at 0, so instrumentation is not required. In what follows we shall focus on the simple OLS, because the estimated parameters of interest do not differ very much after instrumenting. The other IV-estimates are available upon request. In the subsequent subsections, we validate our results on alternative data structures to address the concern of unobserved heterogeneity. We

first discuss our panel analysis and then turn to the dataset of matched pairs of apps.

Panel: We use our monthly panel data to include an app fixed effect (α_i) and to base our estimation on within variation. We thus address the concern that we cannot observe all heterogeneity between apps in the cross sectional analysis, despite our rich set of control variables. For example, while we observe code length and ratings, we cannot evaluate what functionalities two apps provide, how efficiently they perform a given service, or how fun a game really is. Especially unmeasured quality and functionality of apps could be positively correlated with both, app demand and permission usage (resulting in upward bias). To address this concern we estimate the following Fixed-Effects specification:

$$\Delta Demand_{it+1} = \alpha_i + \beta D_{it} + \gamma P_{it} + \theta X_{it} + \varepsilon_{it}. \tag{6}$$

where $\Delta Demand_{it+1}$ now measures the monthly change in downloads. The interpretation of the coefficients of interest, β, differs slightly from that of the cross-section results. A negative significant coefficient would indicate that an increase in permissions use comes with a subsequent decrease in demand growth.

Googles demand measure is too rough to infer monthly downloads. Hence, we have to approximate demand with finer measures that ensure sufficient variance over time. We used three alternative approximations of demand to ensure robustness. First, we use the predicted number of installations (using the number of ratings as an predictor). Alternatively demand was also approximated by the number of ratings directly. Second, we ran a specification with a developer fixed effect. In our third approach we restricted our sample to apps without any change in the length of description over the five months we observed. Apps that do not change their descriptions, more likely offered stable functionality during that period of time.

App Pairs: Our second main strategy to deal with unobserved heterogeneity exploits pairs of apps, which vary only in their price and the amount of permissions. App developers often offer two versions of the same app, one version which can be downloaded for free and one version for which users have to pay a price before installation. These app-pairs shed light on the "money for privacy" trade-off, as it is perceived by developers. The costly version typically offers some advantage over the free version: It may offer additional functionality, and/or contain less advertisement, and/or be associated with fewer (privacy sensitive) permissions. Importantly the paid version serves as technological reference, since any permission that "for pay" does not require is not necessary for the functionality of the app.[16]

We exploit the variation within app pairs to identify the role of permissions in an alternative approach to our panel data analysis. This framework is valid if we can ensure that the two apps within a pair are

[16]If there is any difference paid apps provide *more* functionality.

identical in functionality. For that, we needed to identify app pairs that had no discernible difference in functionality, and only differed in permissions. We manually collected app pairs which stated no difference in their app descriptions, or, as only difference, stated that the free version uses advertisements.The differences in permissions and prices within a pair were used to predict the differences in installations.

6 Results

We now turn to our findings. First we analyze whether cheaper apps use more privacy-intrusive permissions. Second, we present our results on the demand side. Third we analyze circumstantial factors, such as the reputation of app developers, and how they mitigate the strength of the baseline relationship. We present our robustness checks alongside the main results in each subsection.

6.1 Money vs. Privacy on the Supply Side

Table 5 shows descriptive regressions that relate the supply side's pricing choices to the use of privacy sensitive permissions. The two outcomes of interest are the app's business model (free vs. paid; Columns 1-4) and the price charged given it was positive; in Columns 5-8). Columns (1-4) analyze the developer's decision to offer their app for money or for free. The dependent variable is a dummy, which takes the value of 1 if users have to pay a price for downloading the app. Columns 1 and 2 analyze the cross section, while columns 3 and 4 run panel regression. In the cross section we find that apps which use more privacy sensitive permission are between 3.5 and 14.7 percent more likely to be free. Thus, the correlational results confirm the descriptive evidence from above that "A price comes with fewer privacy sensitive permissions." This applies to both, privacy sensitive permissions in general and category-specific sensitive permissions. Secondly the panel regressions in columns 3 and 4 highlight two patterns. First, less than one in 1000 applications ever switch their business model over the period of 30 weeks. Such a low incidence of switching suggests, above all, that the choice of the pricing model is infrequently revoked, and only by a selected subsample in the data. Second, if ever such a switch occurs though, we find that moving from paid to free coincides with an increase in the number of permissions and specifically in add-related permissions.[17] Lastly, as expected, we see that Internet access and Ad-relevant permissions are indeed more likely in free apps.

In columns 5-8 we analyze the logarithm of the price of for paid apps. Columns 5 and 6 show cross section results, while columns 7 and 8 analyze the panel. First, note that only 85,000 (36.6%) have a price, and of those, only 1510 apps (less than 2%) had any variation in price over the six months period of observation. Like for the business model, also the price of an app is hardly ever adjusted. In the cross section (Columns 5 and 6) we see that more privacy sensitive permissions are correlated with higher prices. This unexpected sign highlights that permissions could be correlated with important confounding factors like functionality, service-quality or performance. This unobserved heterogeneity can be controlled in the panel estimation and indeed, in the panel, we see a negative relationship. While the scope of the panel estimation is limited by the

[17] The negative coefficient's literal interpretation is that moving from *free to paid*, reduces the likelihood of observing add related permissions by 49% and the number of permissions by 0.06.

small number of apps we can include, this negative relationship is in line with the notion that apps might be trading more permissions for a lower price (and vice-versa).

Taking all supply specifications together, we conclude that developers are trading access to privacy sensitive information for money. Especially for pay strategies are associated with more privacy for the users, since for pay apps request fewer sensitive permissions and fewer category-specific sensitive permissions.

- Table 5 approximately here

6.2 Demand Side Analysis

Baseline-Specifications and IV Table 4 shows descriptive regressions analyzing the relationship of app downloads and the presence of privacy sensitive permissions. The dependent variable is log(installations). Columns 1-5 show cross section results, with the first column looking at the raw correlation of permissions and downloads. Absent any control variables, the coefficients of privacy sensitive permissions have a positive sign which is presumably related to confounding factors. More permissions could, for example, be related to greater functionality which then leads to more downloads. Indeed, once we introduce control variables (Column 2), the positive coefficient becomes insignificant. In Column 3, we dig one level deeper and consider whether a permission could be required for the service of the app, such as gps location for a runners app. We identify privacy sensitive permissions which are atypical for a given category, and we find that category-specific sensitive permissions are negatively associated with downloads. Similarly, when we account for the number of privacy sensitive permissions as a measure of intensity (Column 4) we also find a negative coefficient.[18] Each additional privacy sensitive permission would be associated with 10 % fewer downloads. In column 5 we differentiate between privacy sensitive permission according to their "functionality" for the developer. The overall picture confirms the findings in columns 3 and 4, and we find that permissions associated to a user's location and communication seem to be more sensitive.

Column 6 and 7 probe into the robustness of these results when accounting for the endogeneity of prices or the unobserved heterogeneity of apps. Column 6 shows a 2SLS instrumental variables estimation to account for the endogeneity of the price choices. The goal of this specification is to test whether the price's endogeneity affects our estimated coefficients for permissions. For the price coefficient we expect an upward bias, as better apps (with more downloads) would charge a higher price. The instruments we use to remedy this problem are the app's code length and the price of the average app's competitors. Indeed, the coefficient of price is more negative in the IV specification, but the coefficients of privacy sensitive permissions are not affected by instrumenting price.

[18] This result is confirmed if we separately estimate indicator variables for the number of permissions (x=1, x=2, ...,x=6,x>6)., with 0 permissions as the reference.

In the last column of the table (7), we analyzes a fixed-effects panel regression for those apps that changed the number of permissions during our period of observation. This strategy addresses our concern that unobserved heterogeneity could bias our cross-sectional estimates, especially the concern of unmeasured quality. The results show that permissions which allow profiling, determining a users' ID or accessing their communication are associated with fewer new downloads, and so is the introduction of ads (unlike in the cross section results where ads (and the related permissions) are associated with greater success). Further reducing the role of unobserved heterogeneity is the main theme of the next table. (6)

- Table 4 approximately here

Panel Data Analysis and App-Pairs In Table 6 we address the concern of unobserved heterogeneity in depth. To do so we use two strategies: In columns 1 to 6 we analyze fixed-effect panel regressions, and in columns 7-9 we use a matched data set which consists of carefully selected app pairs. Columns 1-4 show app-level fixed-effect regressions for apps that changed permissions at least once between April and October 2012. The dependent variable is the approximated number of downloads.[19] We distinguish between the full panel (columns 1 and 2) and the panel that consists only of apps that changed the permissions without a changing the (length) of the app description (which would be expected if the app improved functionality). In Columns 5 and 6 we analyze developer-level fixed effects regressions based on the cross-section data. The concern that motivates this alternative specification is the reduced number of observations when using app-fixed effects, which might introduce selection. To address this issue we exploit the developer dimension in the data by including apps of developers that had ten or more apps in 2012 and adding a fixed effect. This procedure covers more than 50% of our original data set.

Adding privacy sensitive permissions comes with a small negative effect in new downloads (Columns 1 and 3). However, category-specific sensitive permissions (columns 2 and 4) have an almost twice as large negative coefficient, if they come without any updated functionalities (the coefficient is insignificant, on all apps though). Analyzing the regressions with the developer fixed effect (in columns 5 and 6) the picture is confirmed for privacy sensitive permissions. In fact we find a stronger effect than before, but also the standard deviation is increased as the scope of the fixed effect is widened across the potentially quite different apps that a developer might have produced. For category specific sensitive permissions we find the same point estimate as in the panel, but the coefficient is no longer significant.

[19]The approximation based on the categorical variable and the number of new reviews. We use non-linear prediction methods based on the time it takes apps to jump to a higher category and the associated dynamic in the reviews. While our procedure is of great importance for the validity of our panel, it is somewhat involved. Hence, we decided to move more details about this approximation in the data section and an even more detailed data-appendix.

In columns 7-9 of Table 6 we present the results from analyzing closely matched pairs of apps, which is our second strategy to tackle unobserved heterogeneity. Specifically, we use pairs that are offered by the same developer and under the same name, once for free and once for money.[20] App-pairs are interesting, because introducing sensitive permissions in the free but not in the paid version is the most explicit instance of a "money for privacy" trade-off.[21] The second purpose of using app-pairs is applying a second strategy for reducing unobserved heterogeneity. App-pairs are most homogeneous in the sense that they are the same app (name, appearance, developer, etc.), but one is free the other for pay. The cost of the increased homogeneity in app pairs is their scarcity, as we loose many observations, and hence statistical power.

Despite the reduced power in this data set, privacy sensitive permissions are never positive for app pairs (but also never significant). Non-sensitive permissions are positively correlated for all pairs, but the effect goes away, as we use more and more homogeneous pairs. The detailed results are shown in columns 7, 8 and 9 of Table 6. Since the paid "twin" is often a premium version, we expect confounded results when looking at all app pairs , but when keeping functionality increasingly constant, the coefficents should become larger. Indeed, the point estimate for all app pairs (column 7) is very small for sensitive permissions and we even find a positive effect for permissions in general. In column 8 we further restrict the data to only pairs where code-size and the description are the same for both apps (or longer for the free app): This results in a larger (but still insignificant) effect of privacy sensitive permissions. Finally we use only hand-picked pairs, where the paid version does not offer any additional functionality. In this data the positive effect of permissions vanishes, and the point estimate for privacy sensitve permissions remains the same as before (10 percent reduction in demand). These point estimates are in line with the cross-section and panel estimations.

Altogether this section documents that users avoid privacy sensitive permissions once we hold fixed the quality/functionality of an app. Across the two approaches presented in Table 6, panel estimation and the matched data set, we consistently find that downloads are negatively affected by privacy sensitive permissions. Moreover, the effect is consistent across different fixed effect specifications and becomes stronger as we constrain the apps in the data to be increasingly similar. In the next section we analyze which circumstantial factors drive or weaken this result.

- Table 6 approximately here

[20] A typical, though not ideal, pair are a "lite" or "demo" version (free) and the full version (for pay) of the same app, where the free version collects more data. Please refer to the data section for a more detailed description.

[21] The paid version should offer similar or better functionality. Permissions used only in the free version cannot reflect *service*.

6.3 Underlying Mechanisms of the Money vs. Privacy Trade-off

In this section we shed light on the drivers and underlying mechanisms of the money-for-privacy trade-off. We first analyze long-run market outcomes and alternative success measures. Next we focus on circumstantial factors which drive or weaken the negative relationship of permissions and downloads that we presented in the last subsection.

Long-run and Alternative Market Outcomes: Table 10 shows that the results for downloads carry over to alternative and long run performance measures, such as user satisfaction and long term growth. Specifically we analyze an app's growth and survival (Columns 1 to 4), and an app's reputation measured by the the average rating and the number of such ratings by 2012 (Columns 5 to 8). Over this 2-year horizon, app survival (Columns 1 and 2) is 8% less likely if category-atypical sensitive permissions are present (4% for sensitive permissions in general). Similarly, conditional on survival, 2-year-growth of potentially intrusive apps is slower (Columns 3 and 4). Columns 5-8 focus on reviews, which are both an indicator of quality (the average rating), but the total number of reviews also reflects more active app-*usage* (rating an app presupposes using it). We find that ratings are fewer and lower, if an app requires access to sensitive permissions. Especially for category-atypical permissions we find a very strong effect on the number of reviews, which indicates that the negative relationship between permissions and downloads carries over to lower usage.

- Table 10 approximately here

Mitigating Factors: In Table 9 we analyze which factors moderate the relationship of privacy sensitive permissions and installations. The role of privacy concerns should depend on the context, such as the app's category, or its suitability for children, etc. We verify this by contrasting contexts where privacy should matter less (columns 1-3) and contexts where the effect should be larger (4, 5 and 6). In column 7 we analyze how the effect varies by category. In these specifications we include a dummy for for a specific type of app (a special policy, or the app being backed by a brand) *and* a crossterm that estimates the coefficient for requesting privacy sensitive permissions for this special group separately. The dummy is included to account for the mere fact that the selected type app might be systematically different. The cross term captures how installations differ with the presence or absence of privacy sensitive permissions in that group.[22] This analysis serves a double function. First, if privacy concerns drive our results, the relationship of interest should vary by context as theory would predict. Second, we can generate additional insight into how and in which contexts privacy concerns matter in the market for mobile applications.

[22]We thank Catherine Tucker for suggesting this specification, which led to this table.

In columns 1-3 we analyze factors that could reduce how much users worry about their privacy, such as the reputation of the organization behind the app. Column 1 estimates free and paid apps jointly to analyze whether the effect of additional permissions is different for prized apps. In paid apps, the effect of sensitive permissions is indeed much smaller, which could be because users have more trust when the developer received their payment.[23] Column 2 estimates how the relationship of interest differs for apps that have a privacy policy. Apps that have a privacy policy are explicit about the permissions they use and what they will use them for. This would be expected to inspire additional trust from app users. Additional permissions actually appear to have a positive effect when developers state their privacy policy.[24] In Column 3 we seperately analyze corporate apps or apps that belong to large and well known websites (facebook, Amtrak, Banks, Starbucks,...). The reputation of a large corporation should serve as a mechanism to overcome trust-related privacy concerns. To identify such apps, we use additional data from alexa.com to identify apps that are connected to a widely used website.[25] Such apps are typically either corporate apps or they are a mobile spinoff of that site. When analyzing these apps of well known providers seperately, we do not find any penalty for using sensitive permissions. Instead, the sign of the coefficient is completely overturned and the presence of sensitive permissions is associated with a largely increased number of downloads.[26] For large companies that launch an app, privacy sensitive permissions seem to have no negative consequences whatsoever.

In columns 4 and 5 we present robustness checks which are based on our alternative classifications which we obtained from Google's flagging system and from hiring classifiers on Amazon Mechanical Turk. The findings here are largely in line with the main findings: Column 4 analyzes whether the number of apps that are explicitly flagged by Google ("potentially malicious permissions"). This flag is associated with greater visibility, but possibly also a greater potential intrusiveness. Apps that use such permissions are penalized more strongly, and we find much fewer downloads. Column 5 presents the results from using the alternative classification of permission, which we obtained through hiring classifiers on Amazon Mechanical Turk. Based on the classifications provided by the classifiers, we distinguish unproblematic (reference group), somewhat problematic (SP) and very problematic permissions (VP). While apps with somewhat problematic permissions are not installed less frequently than "completely unproblematic apps," very problematic permissions are associated with fewer installations (-13%). Extremely problematic (EP) permissions are the three permissions, which were most frequently classified as problematic. Their presence is associated with even fewer (-31%)

[23] Alternatively this could be different users who selected into buying an app which implies passing on their credit card data.

[24] Note that we do not claim these coefficients to measure the cuasal relationship. We point out, however, that we control for the mere fact that apps with privacy policies might be published by more experienced suppliers (possible better apps). The cross term captures exclusively how users react to the presence or absence of permissions.

[25] We used alexa.com to generate a dummy which takes the value 1 if we could match an app with a highly ranked website.

[26] This result also holds when separately estimating the *number* of sensitive permissions, rather than the dummy. These results are not shown, but are available upon request.

installations on average.

Columns 6-7 we show robustness checks which analyze contexts where privacy concerns might be of varying importance. Column 6 looks at how the coefficient estimates differ for apps that are not suitable for children or young adults. Privacy might be a greater concern in such apps, since they might have different type of content. Users of such apps might prefer not to be identified or might be reluctant to share their contacts with the app developer. The findings shown in column 6 highlight that users of apps with higher maturity level (or no indication) are apparently more privacy sensitive, while apps that are suitable for kids may request more sensitive permissions. In column 7 we differentiated between different categories of apps, because we would expect that the relevance of privacy concerns differs across categories. We distinguish Business, Games/Entertainment/Lifestyle Tools and Educational Apps, and the baseline are health related apps. More details about this classification are provided in the data section. When looking at the different categories we see that users of health and education-related apps seem to avoid privacy sensitive permissions more carefully than users of Tools, Entertainment Apps or Games.

- Table 9 approximately here, condensed version.

The results of this subsection highlight two important features of the money-for-privacy trade-off in the market for mobile applications. First, including privacy sensitive permissions in an app is not only negatively related to downloads, but also to growth, survival, ratings and usage intensity. Second, the trade-off is mediated by contextual factors. Most importantly the negative relationship is weaker in paid apps or in apps which state a privacy policy, and it is overturned when the app is backed by reputation from outside the app market. Our results suggest that the reputation of a well established brand can be used to establish trust.

7 Discussion

We document an important role of private information as a second currency on both sides of the market: (1) App developers clearly ask for more and for more privacy-intrusive permissions if they offer an app for free than if they offer it for a (higher) price and (2) We observe fewer downloads for apps which ask for sensitive permissions, other factors being equal. Furthermore, we see that apps which ask for more permissions (in numbers and in severity) also get fewer reviews, and are less likely to survive two years. Combining both findings, our results highlight a money-for-privacy trade-off in the market for mobile applications: developers offer either a low-priced and more privacy-intrusive app and a higher priced less privacy-intrusive app. Consumers choose between these two options (privacy and money) with a certain understanding that there is no really free lunch.

Our results were found on various subsets of data - a full cross-section, a panel data set, a "long panel" that tracks apps two years later, and carefully selected pairs of apps. We use three alternative classifications to measure the privacy intrusiveness of permissions (Google's classification, Sarma et al. (2012), and Amazon Mechanical Turk). The findings persist for different ways to quantifying intrusiveness (Dummy, number of permissions), and we can show the relationship consistently for both the supply of and the demand for apps. We want to stress that the effects on the demand side are relatively small and cannot easily be isolated from functionality. Nevertheless, the negative relationship between demand and privacy intrusive permissions is robust across multiple specifications and classifications.

In our analysis of circumstancial and moderating factors we both confirm users' general negative attitude to privacy-intrusive apps, but show that this finding does not always hold. It depends on context and breaks down when (commercial) apps can "import" trust through a known brand name or through establishing a privacy policy. Moreover, we analyzed how the relationship of interest differs if an app is offered for a positive price, and the relationship of privacy sensitive permissions and downloads is weaker. The negative effects of permissions were also weaker if (1) privacy-intrusive permissions are common in an app's category, or (2) if a consumer is likely to be relatively young.

The negative relationship between demand and permissions can also become stronger. Context and visibility/awareness (easy interpretability) may play a mediating role here. Especially if the permissions are labeled as malicious by Google, but also if the app belongs to more sensitive categories (such as medicine or business apps) sensitive permissions are associated with even fewer downloads. This finding suggests that more sensitive categories might be up against trust problems, that stand in the way of the provision and usage of very useful services.

Comparing our findings to survey based results or to experimental research in the lab, the effects we see

are clearly smaller CITATION NEEDED(Savage and Waldman (2014) or ??). A simple explanation for this discrepancy would be framing effects or a behaviorally motivated preference to report caring about privacy, these explanations might be too easy. The patterns we see would also be in line with users, who care but do not fully understand. This would reinforce the finding by Savage and Waldman (2014), who show that consumers valued an apps acccess to private information much more negatively than in our paper, but only *after* they received information and training about the permissions' technological implications. We see that Google's warning mechanism seems to result in different behavior, which indicates that simple pieces of information might already suffice to reduce the gap between lab findings and daily life a lot. One possibility to explore this route further would be introducing an easily interpretable traffic light index to indicate how intrusive an app could potentially behave.

Our analysis of mitigating factors highlights that certain (groups of) users might behave quite differently from others. We find different patterns across types of apps and by maturity levels that an app required. This is in line with previous research.CITATION NEEDED (finde nichts wirklich passendes, vl. privacy concerns and expectations are context-dependent, Nissenbaum, 2004 oder Acquisti et al. 2013). The most important mitigating factor to us appears to be the positive interaction of outside reputation and the permissions that app-developers can ask for. Importantly, this is in line with the behavior of consumers who care but do not fully understand, and hence prefer a known brand and distrust an unknown app. On the positive side, allowing established firms to further improve their products based on consumer data can improve quality (but also foster price discrimination). On the negative side, users' reliance on outside reputation would also imply a significant barrier to entry: established brands could promote a new app and gather information about their users much more easily than newcomers. Such a barrier to entry could lead to reduced and a perhaps less than optimal innovative performance of the app market.

Our work suffers from several limitations which highlight avenues for further research. While we do our best to scrutinize the robustnes of our main results, we do not have the ideal data, which would require access to information on downloads and deep app characteristics on performance of an app. Lacking this information, we have to approximate these variables which always involves compromises and insufficiencies, such as a very small dataset and mostly insignificant coefficients on the most closely matched dataset. Moreover, we are aware that the mitigating factors we analyze do not change over time and hence do not afford a panel data analysis. Hence, these additional results are based only on the cross section and represent conditional correlations. For example, the weaker relationship for relatively young consumers needs to be validated with individual usage data. Without such information it is hard to decide, whether young users install apps too lightheartedly or whether older users show too high levels of distrust.

Despite the weaknesses in our data, we see our findings as a first and important step in understanding the role of privacy in app markets. If nothing else, we are at least able to document urgently needed stylized facts. Any policy implications that we suggest should be validated with individual level or experimental data. Great care is warranted, because the app market is a potentially quite sensitive two-sided market which has seen an impressive overall performance. Hence, much of a very good thing is at stake and it is necessary to carefully evaluate any changes in how easily users can evaluate the potential intrusiveness of an app before putting them into practice. Such a careful evaluation would be a fruitful and, we believe, important avenue for further research, since there might be significant unleveraged potentials to further improve the performance of the market especially in sensitive categories.

8 Conclusion

In this paper, we analyze the role of privacy as a second currency in the market for mobile applications. We contribute empirical evidence on a money-for-privacy trade-off between the supply and demand side in this market. We analyzed information for nearly all products which were available in Google's Android Market in 2012 (around 300,000 apps, repeatedly collected in 2012 and 2014). We combined this information with information from additional data sources such as Alexa.com and data from classifiers we hired through Amazon Mechanical Turk. With these data we are able to study the role of privacy for both supply and demand of mobile applications. Specifically, we study (1) whether developers use app permissions to collect private information about users (such as information about their communication behavior, location and profile), and (2) how consumers' installation behavior is related to the presence of privacy-intrusive permissions.

We document - on several data sets, using alternative measures of intrusiveness and multiple specifications - that private information plays an important role on both sides of the market: Developers offer either a lower price for a more privacy intrusive app or a higher priced less privacy-intrusive app. Consumers choose between these two options (privacy and money) with a certain understanding that there is no truly free lunch: We find a weakly negative relationship of permissions and installations across a large number of specifications. We move on to provide nuanced insights on how circumstantial factors mediate the sensitivity of consumers towards apps' permissions: Demand is lower for apps with flagged permissions (by Google), when apps require higher maturity or if they belong to more sensitive categories. Installations and permissions are *not* negatively correlated when the app is associated to websites or developers that have a good reputation. This differential behavior would be in line with consumers, who find it difficult to evaluate whether they are installing a privacy endangering product, and react by favoring brands and apps that they already heard about.

Our results suggest that the app market is a favorable environment for established brands that would like to

develop an app, independently of whether their goal is product improvement or for collecting information about consumers. The flip side interpretation of this result is that the resulting lack of trust towards unknown app developers might constitute a significant barrier to entry which could point to a less than optimal innovative performance of the market for mobile apps especially in "serious" categories, such as health or business. If consumers indeed lack the knowledge to rely on their judgement regarding privacy sensitive permissions, a simple certification or label by an outside party (e.g. a traffic light scheme) could possibly help to reduce this entry barrier.

References

Acquisti, Alessandro and Hal R Varian, "Conditioning Prices on Purchase History," *Marketing Science*, 2005, *24* (3), 367–381.

_ , Curtis R Taylor, and Liad Wagman, "The Economics of Privacy," *Available at SSRN 2580411*, 2015.

_ , Leslie K John, and George Loewenstein, "What is Privacy Worth?," *The Journal of Legal Studies*, 2013, *42* (2), 249–274.

Annie, App, "App Annie Mobile App Forecast: The Path to $100 Billion," 2016.

AppBrain, "Google Play Stats," 2016.

Armstrong, Mark, "Competition in Two-Sided Markets," *The Rand Journal of Economics*, 2006, *37* (3), 668–691.

Askalidis, Georgios, "The Impact of Large Scale Promotions on the Sales and Ratings of Mobile Apps: Evidence from Apple's App Store," Technical Report 2015.

Aziz, Arslan and Rahul Telang, "What is a Cookie Worth?," *Heinz College Working Paper*, 2015.

Carare, Octavian, "The Impact of Bestseller Rank on Demand: Evidence from the App Market*," *International Economic Review*, 2012, *53* (3), 717–742.

Casadesus-Masanell, Ramon and Andres Hervas-Drane, "Competing With Privacy," *Management Science*, 2015, *61* (1), 229–246.

Chevalier, Judith A and Dina Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 2006, *43* (3), 345–354.

Chia, Pern Hui, Yusuke Yamamoto, and N Asokan, "Is this App Safe?: A Large Scale Study on Application Permissions and Risk Signals," in "Proceedings of the 21st international conference on World Wide Web" ACM 2012, pp. 311–320.

Conitzer, Vincent, Curtis R Taylor, and Liad Wagman, "Hide and Seek: Costly Consumer Privacy in a Market with Repeat Purchases," *Marketing Science*, 2012, *31* (2), 277–292.

Davis, Jason P, Yulia Muzyrya, and Pai-Ling Yin, "Experimentation Strategies and Entrepreneurial Innovation: Inherited Market Differences in the iPhone Ecosystem," 2014.

Egelman, Serge, Adrienne Porter Felt, and David Wagner, "Choice Architecture and Smartphone Privacy: There is a Price for That," in "The Economics of Information Security and Privacy," Springer, 2013, pp. 211–236.

Fahl, Sascha, Marian Harbach, Thomas Muders, Lars Baumgärtner, Bernd Freisleben, and Matthew Smith, "Why Eve and Mallory Love Android: An Analysis of Android SSL (in) Security," in "Proceedings of the 2012 ACM Conference on Computer and Communications Security" ACM 2012, pp. 50–61.

Garg, Rajiv and Rahul Telang, "Inferring App Demand from Publicly Available Data," *MIS Quarterly*, 2013, *37* (4), 1253–1264.

Ghose, Anindya and Sang Pil Han, "Estimating Demand for Mobile Applications in the New Economy," *Management Science*, 2014, *60* (6), 1470–1488.

Goldfarb, Avi and Catherine E Tucker, "Privacy Regulation and Online Advertising," *Management Science*, 2011, *57* (1), 57–71.

_ **and Catherine Tucker**, "Shifts in Privacy Concerns," *American Economic Review: Papers and Proceedings*, 2012, *102* (3), 349–353.

Gross, Ralph and Alessandro Acquisti, "Information Revelation and Privacy in Online Social Networks," in "Proceedings of the 2005 ACM workshop on Privacy in the Electronic Society" ACM 2005, pp. 71–80.

Grossklags, Jens and Alessandro Acquisti, "When 25 Cents is Too Much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information," in "Workshop on the Economics of Information Securtiy (WEIS)" 2007.

Hann, Il-Horn, Kai-Lung Hui, IPL Png, and Sang-Yong Tom Lee, "Direct Marketing: Privacy and Competition," 2004.

IDC, "Worldwide Quarterly Mobile Phone Tracker," 2015.

Johnson, Garrett A, "The Impact of Privacy Policy on the Auction Market for Online Display Advertising," 2013.

Johnson, Justin P, "Targeted Advertising and Advertising Avoidance," *The Rand Journal of Economics*, 2013, *44* (1), 128–144.

Marthews, Alex and Catherine Tucker, "Government Surveillance and Internet Search Behavior," Technical Report 2014.

OECD, "Broadband database," 2012.

Preibusch, Sören and Joseph Bonneau, "The Privacy Landscape: Product Differentiation on Data Collection," in "Economics of Information Security and Privacy III," Springer, 2013, pp. 263–283.

Rochet, Jean-Charles and Jean Tirole, "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, 2003, *1* (4), 990–1029.

Sarma, Bhaskar Pratim, Ninghui Li, Chris Gates, Rahul Potharaju, Cristina Nita-Rotaru, and Ian Molloy, "Android Permissions: A Perspective Combining Risks and Benefits," in "Proceedings of the 17th ACM symposium on Access Control Models and Technologies" ACM 2012, pp. 13–22.

Savage, Scott J and Donald M Waldman, "The Value of Online Privacy," *Discussion Paper in Economics, University of Colorado at Boulder*, 2013.

Savage, Scott J. and Donald M. Waldman, "The Value of Online Privacy: Evidence from Smartphone Applications," 2014.

Spiegel, Yossi, "Commercial Software, Adware, and Consumer Privacy," *International Journal of Industrial Organization*, 2013, *31* (6), 702–713.

Taylor, Curtis R, "Consumer Privacy and the Market for Customer Information," *The Rand Journal of Economics*, 2004, *35* (4), 631–650.

_ **, Vincent Conitzer, and Liad Wagman**, "Online Privacy and Price Discrimination," *Economic Research*, 2010.

Tsai, Janice Y, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti, "The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study," *Information Systems Research*, 2011, *22* (2), 254–268.

Tucker, Catherine, "The Economics of Advertising and Privacy," *International Journal of Industrial Organization*, 2012, *30* (3), 326–329.

_ , "Social Networks, Personalized Advertising and Privacy Controls," *Journal of Marketing Research*, 2014, *51* (5), 546–562.

Wathieu, Luc, "Privacy, Exposure and Price Discrimination," *Harvard Business School Marketing Research Paper No. 02-03*, 2002.

Yin, Pai-Ling, Jason P Davis, and Yulia Muzyrya, "Entrepreneurial Innovation: Killer Apps in the iPhone Ecosystem," *American Economic Review: Papers and Proceedings*, 2014, *104* (5), 255–259.

A Descriptive Tables

Table 2: Permission Group Definitions

Permissions (Group)	Description	Sarma	Google	MTurk	$D_{PrivCatSpec}$
$D_{Privacy}$					
D_{ID}					
read phone state and ID	Allows read only access to phone state.	1	0	0	2,7,9,11,13,14,19,21,23,30
$D_{Location}$					
coarse location	Allows an app to access approximate location derived from network location sources such as cell towers and Wi-Fi.	1	0	0	1,2,3,5,6,7,9,10,11,16,17,19, 22,26
fine gps location	Allows an app to access precise location from location sources such as GPS, cell towers, and Wi-Fi.	1	1	0	1,2,3,5,6,7,9,10,11,13,15,16, 17,19,22,26
$D_{Communication}$					
intercept outgoing calls	Allows an app to see the number being dialed during an outgoing call with the option to redirect the call to a different number or abort the call altogether.	1	1	1	1,2,3,5,6,7,9,10,11,12,14,16, 18,19,20,22,23,25,26,29,30
read sms or mms	Allows an app to read SMS and MMS messages.	1	1	1	1,2,3,5,6,7,9,11,12,13,14,15, 16,17,18,19,20,22,23,25,26,29,30
receive sms	Allows an app to monitor incoming SMS messages, to record or perform processing on them.	1	1	1	1,2,3,5,6,7,9,10,11,12,13,14, 15,16,17,19,20,22,25,26,30
receive mms	Allows an app to monitor incoming MMS messages, to record or perform processing on them.	1	1	1	1,2,3,4,5,6,7,9,10,11,12,14,15, 16,18,19,20,22,25,26,28,29,30
record audio	Allows an app to record audio.	1	0	1	1,2,3,5,6,7,11,13,14,16,18,19, 22,23,26,30
receive wap	Allows an app to monitor incoming WAP push messages.	1	1	0	1,2,3,5,6,7,9,10,11,12,14,15, 16,18,20,22,25,26,28,29,30
$D_{Profile}$					
read contact data	Allows an app to read the user's contacts data.	1	1	1	1,2,3,5,6,7,9,10,11,12,13,14, 15,16,17,18,19,20,22,26,30
read browser data	Allows an app to read (but not write) the user's browsing history and bookmarks.	1	0	1	1,2,3,5,6,7,9,10,11,12,14,16, 17,19,20,22,24,25,26,29,30
read sensitive log data	Allows an app to read the low-level system log files.	1	0	1	1,2,3,5,6,7,9,10,11,12,13,14, 16,19,20,22,25,26,28

Notes: Numbers in column $D_{PrivCatSpec}$ indicate: 1: Arcade&Action, 2: Books&Reference, 3: Brain&Puzzle, 4: Business, 5: Cards&Casino, 6: Casual, 7: Comics, 8: Communication, 9: Education, 10: Entertainment, 11: Finance, 12: Health&Fitness, 13: Libraries&Demo, 14: Lifestyle, 15: Media&Video, 16: Medical, 17: Music&Audio, 18: News&Magazines, 19: Personalization, 20: Photography, 21: Productivity, 22: Racing, 23: Shopping, 24: Social, 25: Sports, 26; Sports Games, 27: Tools, 28: Transportation, 29: Travel&Local, 30: Weather. The D_i variables are dummy variables which are equal to one if an app uses one of the permissions of a respective permission group. D_{Other} consists of: mount and unmount file systems, add or modify calendar events and send, write contact data, write browser history and bookmark, edit sms or mms, modify delete usb storage contents, control near field communication, view configured accounts, create bluetooth connections, bluetooth administration, directly call any phone numbers, send sms messages. $D_{Internet}$ consists of permissions full internet access and D_{Ads} consists of permission view network state.

App Categories: Finally we used Google's categorization to identify seven overarching categories of apps. Table 3 shows how we classified Android's thirty categories into seven overarching meta categories. This was merely done for simplified representation for category-specific results.[27] The table shows all the 30 categories in Android's Playstore sorted by their size. Moreover it indicates to which of our seven overarching categories the categories were added. We defined Education, Entertainment, Games, Tools&Personalization, Lifestyle, Health and Business. Surprisingly, in 2012, Games were not the largest category of apps. Instead, the largest resulting meta category is Tools&Personalization (69,372 apps), which also contain weather and transportation apps. The smallest are Health and Business related apps (8,255 and 11,686 apps respectively).

Table 3: Classification of Categories into seven Meta Categories.

	b	pct	
Personalization	28819	12.3	Tools & Personalization
Entertainment	25179	10.8	Entertainment
Tools	20558	8.8	Tools & Personalization
Books & Reference	15624	6.7	Education
Brain & Puzzle	14687	6.3	Games
Lifestyle	12238	5.2	Lifestyle
Education	11314	4.8	Education
Travel & Local	10419	4.5	Lifestyle
Arcade & Action	8308	3.6	Games
Productivity	8215	3.5	Tools & Personalization
Casual	7963	3.4	Games
Music & Audio	7907	3.4	Entertainment
Sports	7715	3.3	Lifestyle
Business	7239	3.1	Business
Communication	5874	2.5	Tools & Personalization
Health & Fitness	5551	2.4	Health
News & Magazines	5160	2.2	Entertainment
Social	4926	2.1	Entertainment
Finance	4447	1.9	Business
Media & Video	3572	1.5	Entertainment
Photography	2769	1.2	Tools & Personalization
Medical	2704	1.2	Health
Shopping	2525	1.1	Lifestyle
Transportation	2150	0.9	Tools & Personalization
Cards & Casino	2105	0.9	Games
Sports Games	1446	0.6	Games
Comics	1413	0.6	Entertainment
Libraries & Demo	1301	0.6	Education
Weather	987	0.4	Tools & Personalization
Racing	698	0.3	Games
Observations	233813		

[27]Category-specific privacy sensitive permissions are computed based on the 30 underlying categories. The category-specific results do not depend on this classification, the coefficients, for all 30 categories estimated separately are available upon request.

B Estimation Tables

Table 4: Baseline Demand Side Results

	Cross-Section					IV-Paid	Panel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\#_{TotalPerm.}$	0.090***	0.010***	0.019***	0.037***	0.033***	0.057***	-0.006***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.009)	(0.001)
$D_{Privacy}$	0.261***	-0.014					-0.024**
	(0.018)	(0.012)					(0.011)
$D_{PrivCatSpec}$			-0.252***				
			(0.014)				
$\#_{PrivacyPerm.}$				-0.103***			
				(0.007)			
$D_{Location}$					-0.274***	0.083*	
					(0.018)	(0.044)	
$D_{Communication}$					-0.178***	-0.143***	
					(0.020)	(0.042)	
$D_{Profile}$					-0.069***	-0.119**	
					(0.016)	(0.051)	
D_{ID}					-0.078***	-0.113***	
					(0.013)	(0.034)	
$D_{Internet}$		-0.014	-0.005	-0.030**	-0.018	0.097***	0.022
		(0.012)	(0.012)	(0.012)	(0.012)	(0.024)	(0.021)
D_{Ads}		0.105***	0.098***	0.076***	0.073***	-0.124***	-0.012
		(0.011)	(0.011)	(0.012)	(0.012)	(0.028)	(0.011)
D_{Other}		0.039***	0.037***	0.018*	0.025**	0.138***	-0.012
		(0.011)	(0.011)	(0.011)	(0.011)	(0.039)	(0.010)
D_{Price}		-2.361***	-2.411***	-2.370***	-2.382***		2.061***
		(0.123)	(0.123)	(0.123)	(0.123)		(0.654)
Log. Price		-0.066***	-0.063***	-0.067***	-0.066***	-0.376**	-0.085
		(0.010)	(0.010)	(0.010)	(0.010)	(0.149)	(0.056)
Constant	0.869***	2.987***	2.946***	2.902***	2.853***	1.685***	-0.179
	(0.032)	(0.213)	(0.213)	(0.213)	(0.213)	(0.346)	(0.663)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	233813	233813	233813	233813	233813	85569	28145
Num. of Groups							7176
Mean of dep. Var.	-0.66	-0.66	-0.66	-0.66	-0.66	-3.35	0.14
SD of dep. Var.	3.49	3.49	3.49	3.49	3.49	2.86	0.23
Adjusted R^2	0.448	0.691	0.692	0.692	0.692	0.496	0.078

Notes: The table shows descriptive regressions analyzing the relationship of app downloads and the presence of privacy sensitive permissions. Columns 1-5 show cross section results, Column 6 analyzes panel data and in Column 7 we show 2SLS instrumental variables estimation to account for the endogeneity of the price choices. Column 1 shows the raw correlation between privacy sensitive permissions and downloads, without controlling for app performance. In column 2 we introduce our control variables to reduce unobserved heterogeneity. In column 3 we look at privacy sensitive permission that are not commonly used in the app's category. In Column 4 we introduce the number of privacy sensitive permissions and column 5, finally, disaggregates the privacy sensitive permissions into functionality related types of permissions. Column 6 shows fixed effects panel regressions that use only variation within a given app. Column 7, lastly repeats the specification in column 2, but instruments for price, to account for the likely endogeneity of this variable. The coefficient of interest analyzes the relationship between an app's downloads and our measures of privacy sensitive permissions. These measures are category-specifically atypical sensitive permission and the broader measure of generally privacy sensitive permissions. Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Baseline Supply Side Results

	Business Model Choice (D_{Price})				Price Choice (Log. Price)			
	CS	CS	Panel	Panel	CS	CS	Panel	Panel
$\#_{TotalPerm.}$	-0.004***	0.000	-0.062**	-0.051**	0.017***	0.023***	0.019***	0.019***
	(0.000)	(0.000)	(0.026)	(0.022)	(0.002)	(0.002)	(0.006)	(0.006)
$D_{Privacy}$	-0.035***		0.121		0.118***		-0.101***	
	(0.002)		(0.082)		(0.009)		(0.029)	
$D_{PrivCatSpec}$		-0.147***		-0.022		0.072***		-0.168***
		(0.003)		(0.084)		(0.012)		(0.040)
$D_{Internet}$	-0.192***	-0.187***	-0.073	-0.042	0.081***	0.084***	0.149***	0.146***
	(0.003)	(0.003)	(0.078)	(0.078)	(0.007)	(0.007)	(0.029)	(0.030)
D_{Ads}	-0.127***	-0.129***	-0.497***	-0.490***	-0.015*	-0.017**	-0.030	-0.035
	(0.002)	(0.002)	(0.062)	(0.059)	(0.008)	(0.008)	(0.030)	(0.029)
D_{Other}	0.059***	0.056***	0.099	0.124	0.218***	0.219***	0.053*	0.045
	(0.002)	(0.002)	(0.083)	(0.100)	(0.007)	(0.007)	(0.030)	(0.030)
Constant	-0.038	-0.076**	3.166***	3.376***	-0.030	-0.037	1.011*	1.070*
	(0.036)	(0.036)	(0.914)	(0.923)	(0.108)	(0.108)	(0.571)	(0.578)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	233813	233813	1145	1145	85569	85569	7002	7002
Num. of Groups			229	229			1510	1510
Mean of dep. Var.	0.37	0.37	0.52	0.52	0.27	0.27	0.68	0.68
SD of dep. Var.	0.48	0.48	0.50	0.50	0.77	0.77	0.85	0.85
Adjusted R^2	0.398	0.405	0.653	0.651	0.271	0.269	0.047	0.049

Notes: The table shows descriptive regressions analyzing the pricing choices of the supply Side. In Columns 1-4 the dependent variable is the developer's decision to offer their app for money or for free (Dummy, taking the value 1 if free). In Columns 5-8 the dependent variable is the price of an app, given that the developer chose the model. In each block, the first two columns analyze the cross section of all apps, while the second two columns (3,4, 7 and 8) analyze panel data for the restricted set of apps that changed their policy in our period of observation. The coefficient of interest analyzes the relationship between an apps pricing policy and our two main measures of an apps use of privacy sensitive permissions. These measures are category-specifically atypical sensitive permission (even columns) and the broader measure of generally privacy sensitive permissions (odd columns). Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Panel Demand Side Results

	Panel(Pred.Installs.)		Restr.Panel(Pred.Installs.)		CS (Dev FE, Installs.)		Pairs (FE, Installs.)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\#_{TotalPerm.}$	-0.009***	-0.009***	-0.006***	-0.006***	0.020***	0.016***	0.052***	0.068*	-0.017
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.015)	(0.037)	(0.060)
$D_{Privacy}$	-0.021**		-0.024**		-0.063**		-0.024	-0.104	-0.098
	(0.008)		(0.011)		(0.027)		(0.062)	(0.113)	(0.189)
$D_{PrivCatSpec}$		-0.012		-0.035***		0.036			
		(0.009)		(0.012)		(0.027)			
$D_{Internet}$	-0.013	-0.014	0.022	0.020	0.039	0.034	0.026	0.017	-0.061
	(0.012)	(0.012)	(0.021)	(0.021)	(0.029)	(0.029)	(0.049)	(0.101)	(0.149)
D_{Ads}	0.001	-0.000	-0.012	-0.012	0.173***	0.172***	0.267***	0.118	0.000
	(0.008)	(0.008)	(0.011)	(0.011)	(0.028)	(0.028)	(0.049)	(0.094)	(0.179)
D_{Other}	-0.014**	-0.014**	-0.012	-0.013	0.024	0.022	0.075	0.166	-0.195
	(0.007)	(0.007)	(0.010)	(0.010)	(0.025)	(0.025)	(0.073)	(0.144)	(0.220)
D_{Price}	2.231***	2.224***	2.061***	2.057***	-1.588***	-1.572***	-6.067***	-7.691***	-4.709***
	(0.308)	(0.308)	(0.654)	(0.652)	(0.271)	(0.271)	(0.336)	(0.637)	(1.619)
Log. Price	-0.102***	-0.101***	-0.085	-0.085	-0.190***	-0.191***	0.166***	0.295***	-0.013
	(0.024)	(0.024)	(0.056)	(0.055)	(0.023)	(0.023)	(0.028)	(0.054)	(0.136)
Constant	-0.482	-0.474	-0.179	-0.163	1.873***	1.852***	0.327	10.800***	0.815
	(0.310)	(0.309)	(0.663)	(0.660)	(0.428)	(0.428)	(1.434)	(3.366)	(2.813)
Category	Yes	Yes	Yes	Yes	No	No	No	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	61927	61927	28145	28145	112414	112414	18780	5336	992
Num. of Groups	15739	15739	7176	7176			9390	2668	496
Mean of dep. Var.	0.15	0.15	0.14	0.14	-1.70	-1.70	-0.02	-0.20	0.03
SD of dep. Var.	0.24	0.24	0.23	0.23	3.64	3.64	3.48	3.47	3.48
Adjusted R^2	0.069	0.069	0.078	0.078	0.801	0.801	0.898	0.902	0.938

Notes: This table shows the results from fixed-effect panel regressions. The dependent variable is log(installations). Columns 1-4 show the panel that relies on apps that changed permissions at least once between April and October 2012. The new downloads are approximated by an estimation procedure based on monthly new reviews. The first two columns show all apps with a permission change, while columns 3-4: show when new permissions were introduced without any significant improvements in functionality (no version update). Columns 5 and 6 analyze cross-section data, but introduce a dummy to control for the developers of the apps if the developer has 10 ore more apps (). Finally in columns 7-9 we analyze which are composed of the same app (name, appearance, developer, etc.), but one version of the pair is for free the other for pay. App pairs are most homogeneous in the sense that they look the same, have the same name and perform the same task, with only a small difference, either in functionality, advertisements or sensitive permissions. Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Alternative Market Outcomes and Success Condensed

	App Survival		Growth		Num of Ratings		User Assessment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\#_{TotalPerm.}$	-0.011***	-0.010***	-0.003*	-0.004***	0.029***	0.038***	0.001**	-0.000
	(0.001)	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)
$D_{Privacy}$	-0.041***		-0.046***		-0.021**		-0.013***	
	(0.003)		(0.009)		(0.010)		(0.002)	
$D_{PrivCatSpec}$		-0.080***		-0.037***		-0.265***		0.001
		(0.003)		(0.010)		(0.012)		(0.002)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	233813	233813	167104	167104	233813	233813	233813	233813
Avg(dep. Var.)	0.71	0.71	0.62	0.62	-0.91	-0.91	1.33	1.33
SD(dep. Var.)	0.45	0.45	1.15	1.15	6.34	6.34	0.29	0.29
Adjusted R^2	0.188	0.190	0.065	0.065	0.946	0.946	0.078	0.078

Notes: This table shows the results for alternative performance measures of mobile apps. Columns 1 to 4 analyze App survival (Cols. 1+2) and App growth (conditional on survival; Cols. 3+4) over a 2-year horizon. Columns 5-8 focus on reviews, and specifically for the number of new reviews (Cols. 5-6) and their average grade (Cols. 7-8). Each specification is estimated for both privacy sensitive permissions in general (odd columns) and category specific sensitive permissions (even). Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

B.1 Paper Appendix

Table 8: Mitigating Factors Condensed

Mitigating factors	Price	Privacy Policy	Alexa.com	Google	AMTurk	Maturity
	(1)	(2)	(3)	(4)	(5)	(6)
No	-0.049***	-0.031**	-0.019	0.086***	0.014	-0.045***
	(0.014)	(0.012)	(0.013)	(0.014)	(0.013)	(0.012)
Yes	0.064**	0.503***	0.811***	-0.073**	<-0.131***	-0.151***
	(0.020)	(0.046)	(0.151)	(0.027)	(0.016)	(0.028)

Analysis by category	Health	Education	Tools	Business	EntGameLife
	(1)	(2)	(3)	(4)	(5)
$Category\ Coeff$	*baseline*	0.188***	0.536***	0.366***	0.459***
$Net\ Effect$	-0.478***	-0.290***	0.058*(?)	-0.112*(?)	-0.019

Notes: This table shows the main results from our analysis of factors that moderate the role of privacy sensitive permissions. The dependent variable is log(installations). Column 1 in the upper panel estimates free and paid apps jointly to analyze whether the effect of additional permissions is different for prized apps. Column 2 estimates how the relationship of interest differs for apps that have a privacy policy, and column 3 separately analyzes apps that are connected to a widely used website (corporate apps and well known websites that have a high ranking on Alexa.com). Column 4 analyzes whether the number of apps that are explicitly flagged by Google are penalized more strongly. Column 5 uses an alternative classification of sensitive permission that was obtained from hiring 400 workers on Amazon's Mechanical Turk. Column 6 looks at how the coefficient estimates differ for apps that are not suitable for children or young adults. The second panel shows the results when differentiating between different categories of apps. We distinguish Business, Games/Entertainment/Lifestyle Tools and Educational Apps. The baseline are health related apps. Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Mitigating Factors

	Price	Privacy Policy	Alexa.com	Google	MTurk	Maturity	Categories
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\#_{TotalPerm.}$	0.011***	0.008***	0.011***	0.022***	0.025***	0.007***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$D_{Privacy}$	-0.049***	-0.031**	-0.019	0.086***		-0.045***	-0.478***
	(0.014)	(0.012)	(0.013)	(0.014)		(0.012)	(0.043)
$D_{Internet}$	-0.021	-0.010	0.010	-0.027**	-0.038***	-0.015	-0.016
	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.012)	(0.012)
D_{Ads}	0.105***	0.099***	0.083***	0.074***	0.081***	0.101***	0.087***
	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)
D_{Other}	0.039***	0.043***	0.049***	0.036***	0.024**	0.045***	0.050***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
D_{Price}	-2.306***	-2.379***	-2.262***	-2.279***	-2.379***	-2.363***	-2.410***
	(0.123)	(0.122)	(0.130)	(0.123)	(0.123)	(0.123)	(0.122)
Log. Price	-0.074***	-0.064***	-0.074***	-0.074***	-0.066***	-0.065***	-0.060***
	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)
$D_{Privacy} \times D_{Price}$	0.113***						
	(0.021)						
$D_{PrivacyPolicy}$		0.302***					
		(0.033)					
$D_{Privacy} \times D_{PrivacyPolicy}$		0.534***					
		(0.046)					
$D_{HighAlexaRank}$			1.176***				
			(0.112)				
$D_{Privacy} \times D_{HighAlexaRank}$			0.830***				
			(0.151)				
$D_{GoogleMalicious}$				-0.129***			
				(0.025)			
$D_{Privacy} \times D_{GoogleMalicious}$				-0.159***			
				(0.029)			
$D_{MTurkSP}$					0.014		
					(0.013)		
$D_{MTurkVP}$					-0.131***		
					(0.016)		
$D_{MTurkEP}$					-0.310***		
					(0.028)		
$D_{HighNoMaturity}$						0.340***	
						(0.017)	
$D_{Privacy} \times D_{HighNoMaturity}$						-0.106***	
						(0.028)	
$D_{Privacy} \times D_{Education}$							0.188***
							(0.049)
$D_{Privacy} \times D_{Tools}$							0.536***
							(0.045)
$D_{Privacy} \times D_{Business}$							0.367***
							(0.055)
$D_{Privacy} \times D_{EntGameLife}$							0.459***
							(0.044)
Constant	2.907***	3.070***	3.003***	2.800***	2.859***	3.107***	3.382***
	(0.214)	(0.213)	(0.228)	(0.214)	(0.214)	(0.213)	(0.213)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	233815	233815	206504	233815	233815	233815	233815
Mean of dep. Var.	-0.66	-0.66	-0.68	-0.66	-0.66	-0.66	-0.66
SD of dep. Var.	3.49	3.49	3.52	3.49	3.49	3.49	3.49
Adjusted R^2	0.691	0.692	0.696	0.692	0.692	0.691	0.691

Notes: This table analyzes factors that moderate the role of privacy sensitive permissions. The dependent variable is log(installations). Column 1 estimates free and paid apps jointly to analyze whether the effect of additional permissions is different for prized apps. Column 2 estimates how the relationship of interest differs for apps that have a privacy policy, and column 3 separately analyzes apps that are

Table 10: Alternative Market Outcomes and Success Measures

	App Survival		Growth		Num of Ratings		User Assessment	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\#_{TotalPerm.}$	-0.011***	-0.010***	-0.003*	-0.004***	0.029***	0.038***	0.001**	-0.000
	(0.001)	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)
$D_{Privacy}$	-0.041***		-0.046***		-0.021**		-0.013***	
	(0.003)		(0.009)		(0.010)		(0.002)	
$D_{PrivCatSpec}$		-0.080***		-0.037***		-0.265***		0.001
		(0.003)		(0.010)		(0.012)		(0.002)
$D_{Internet}$	-0.034***	-0.035***	-0.021***	-0.024***	-0.158***	-0.149***	-0.029***	-0.031***
	(0.003)	(0.002)	(0.008)	(0.008)	(0.009)	(0.009)	(0.002)	(0.002)
D_{Ads}	-0.007***	-0.008***	0.095***	0.095***	0.124***	0.118***	0.012***	0.012***
	(0.002)	(0.002)	(0.007)	(0.007)	(0.009)	(0.009)	(0.002)	(0.002)
D_{Other}	0.023***	0.021***	0.011	0.010	-0.010	-0.013	-0.003**	-0.004**
	(0.002)	(0.002)	(0.008)	(0.008)	(0.009)	(0.009)	(0.001)	(0.001)
D_{Price}	-0.183***	-0.187***	-0.495***	-0.486***	-2.073***	-2.122***	0.086***	0.090***
	(0.024)	(0.024)	(0.081)	(0.081)	(0.074)	(0.074)	(0.017)	(0.017)
Log. Price	0.020***	0.020***	0.016**	0.015**	0.088***	0.091***	-0.010***	-0.011***
	(0.002)	(0.002)	(0.007)	(0.007)	(0.006)	(0.006)	(0.001)	(0.001)
Log. Installations (in 1000)	0.003***	0.003***						
	(0.000)	(0.000)						
Constant	1.478***	1.456***	-1.635***	-1.649***	3.383***	3.338***	0.402***	0.399***
	(0.045)	(0.045)	(0.139)	(0.139)	(0.159)	(0.158)	(0.031)	(0.031)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	233813	233813	167104	167104	233813	233813	233813	233813
Mean of dep. Var.	0.71	0.71	0.62	0.62	-0.91	-0.91	1.33	1.33
SD of dep. Var.	0.45	0.45	1.15	1.15	6.34	6.34	0.29	0.29
Adjusted R^2	0.188	0.190	0.065	0.065	0.946	0.946	0.078	0.078

Notes: This table shows the results for alternative performance measures of mobile apps. Columns 1 to 4 analyze App survival (Cols. 1+2) and App growth (conditional on survival; Cols. 3+4) over a 2-year horizon. Columns 5-8 focus on reviews, and specifically for the number of new reviews (Cols. 5-6) and their average grade (Cols. 7-8). Each specification is estimated for both privacy sensitive permissions in general (odd columns) and category specific sensitive permissions (even). Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

B.2 Figures

Figure 4: App Information in the Android Market 2012

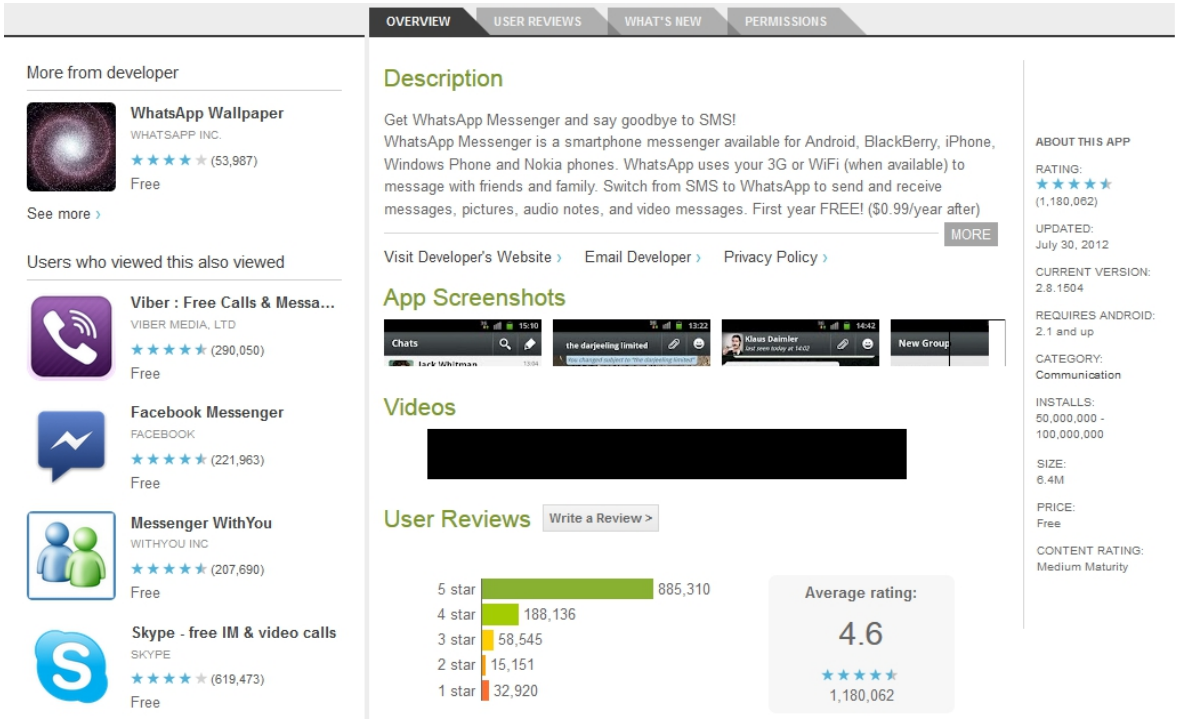


Figure 5: Permission Information in the Android Market 2012