When Private Information Settles the Bill: Money and Privacy in Google's Market for Smartphone Applications

Michael Kummer and Patrick Schulte

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 16-031

When Private Information Settles the Bill: Money and Privacy in Google's Market for Smartphone Applications

Michael Kummer and Patrick Schulte

When Private Information Settles the Bill:

Money and Privacy in Google's Market for Smartphone Applications[*]

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

Patrick Schulte[‡]
Centre for European
Economic Research (ZEW)

This Version: April 2016
First Version: December 2014

Abstract

We shed light on a money-for-privacy trade-off in the market for smartphone applications ("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 these data with information from Alexa.com and Amazon Mechanical Turk. Our findings show that both the market's supply and the demand side consider an app's ability to collect private information, measured by their use of privacy-sensitive permissions: (1) cheaper apps use more privacy-sensitive permissions; (2) installation numbers are lower for apps with sensitive permissions; (3) circumstantial factors, such as the reputation of app developers, mitigate the strength of this relationship. Our results emerge consistently across several robustness checks, including the use of panel data analysis, the use of selected matched "twin"-pairs of apps and the use of various alternative measures of privacy-sensitiveness.

JEL Classification: D12, D22, L15, L86

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

1 Introduction

Striking the balance between "too little" and "too much" privacy protection in new digital tech-
nologies and online markets might be among the biggest challenges of our times. On the one
hand, the market success of new digital technologies may very much depend on the services'
ability to collect and analyze *enough* personal information (Goldfarb and Tucker, 2011; John-
son, 2013b; Aziz and Telang, 2015). The resulting products bear enormous potential for better
information flows, better choices, efficiency and increased welfare. On the other hand, when
providers store *too much* personal data this could result in widespread unease and loss of trust in
the market (Miller and Tucker, 2009; Acquisti et al., 2015, forthcoming).[1] Ultimately, too much
data collection may also imply significant societal risks.[2]

In this paper, we analyze data on 300,000 smartphone applications (henceforth "apps") to
inform the debate about optimal levels of privacy protection. Using these data from Google's
Android Market, we study the extent to which private information has taken the role of a "second
currency" on both sides of the market for apps. We provide the first large scale and market-
transaction based evidence on the money-for-privacy trade-off in this market. App developers
trade greater access to personal information for lower prices and consumers choose between
lower prices and more privacy. The app-market is relevant for our question, because apps have
transformed information exchanges in less than eight years, and they offer unseen potentials
for collecting private user information at low cost. Moreover, app market data are informative,
because we can exploit Google's unique policy of highlighting an app's ability to access private
information of users. Android's Operating System confronts all users with the complete list of
rights (henceforth "permissions") that an app requests. Users must grant these permissions before
installation. Thanks to this policy we could discern 136 distinct permissions in apps and record
the permission requirements of each app in 2012.

We combine this information on each app's ability to collect personal information with a rich
data set on 300,000 apps from Google's Android Market. The data covers all publicly available
app-specific information including each app's number of installation, its price and even each
app's closest competitors. We collected the data repeatedly in 2012 (over 6 months) and once
in 2014. We augmented these data with information from Alexa.com and with data we collected
on Amazon Mechanical Turk to add further background information about app providers and

---

[1] According to the Pew Research Center, 68% of adults believe current laws to insufficiently protect individuals'
online privacy (Rainie et al., 2013).

[2] We refer to real historical experiences e.g. in Europe or Asia, where too much data in the wrong hands
contributed to totalitarianism, persecution, mass murder and war. Recent events (in the United States and
elsewhere) show that governments continue to have strong incentives to access the data collected by new online
technologies.

permissions.

Our results document a money-for-privacy trade-off on both sides of the market: (1) App developers clearly ask for more privacy-sensitive permissions if they offer an app for free than if they offer it for a (higher) price. (2) Consumers take this trade-off into account and show up to 25% lower demand for apps which ask for privacy-sensitive permissions. This inference is based on a full cross-section, a monthly panel, a 2-year panel and a carefully selected set of "twin-pairs" of apps. The results consistently emerge for both the supply and demand side across these different data sets and in various specifications. They also are robust to using various alternative outcomes, to using an IV-approach to instrument for apps' prices, and to using three alternative definitions of intrusiveness: a classification by Sarma et al. (2012), one by Google and one based on classifiers from Amazon Mechanical Turk. We generate additional insight from an analysis of "twin-pairs" of apps. These are two versions of the same app, where one is for free and the other one is for pay, and where the paid version can serve as a technological reference to identify redundant permissions.

Finally, in addition to highlighting the money-for-privacy trade-off, we illuminate which factors mediate its strength. The trade-off appears to be weaker for well known brands than for new or unknown developers. This effect might constitute a significant barrier to entry for new developers which may result in suboptimal innovation in the market for mobile applications. If this result is driven by consumers' lack of knowledge about privacy-sensitive permissions, policy makers might want to consider user-friendly certification procedures by outside parties (e.g. a traffic light scheme) to help reduce this entry barrier (Tsai et al., 2011).

Our paper follows the usual structure. We first review the related literature (Section 2) and provide background information on the app market (Section 3). Next we describe the data set (Section 4) and discuss our estimation strategies for both sides of the market (Section 5). Section 6 presents results and robustness checks. Section 7 contains a discussion and Section 8 concludes.[3]

2 Related Literature and Contribution

The value of private data has become a central theme of the (economic) public debate. Experts had issued warnings about the possibilities of tracking individuals online already for many years. While these were widely ignored before 2013, the news of the U.S. National Security Administration's (NSA) data gathering and analysis has increased public awareness and raised considerable concerns. Hence, a small but growing stream of research provided urgently needed first empirical

---

[3]Appendices contain tables, figures, additional details about the data and further robustness checks.

evidence on the role of private information for supply and demand in online markets.[4] Preibusch and Bonneau (2013) analyze data collection policies of internet sites and find large variation in how intensely they collect user data. A series of related 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 restricting the use of private data in advertisement reduced targeting effectiveness, which resulted in lower revenues for the content site, but also highlighted the important effect of privacy policies on consumer behavior.[5]

On the demand side Gross and Acquisti (2005) study demand for privacy in social networks. Very few users change the highly permeable privacy preferences. Acquisti et al. (2013) use a field experiment to examine consumer preferences for privacy. They link the sensitivity of privacy valuations to contextual factors, like the framing of a webpage. Goldfarb and Tucker (2012) use information from an online survey to find that privacy concerns have increased over time, especially for older people. Marthews and Tucker (2014) study the effect of government surveillance on internet search behavior. When the NSA surveillance became known, Internet users less frequently searched for terms they suspected the US government to track. Turow et al. (2009) find most Americans (66%) disapprove tailored advertisements, and young adults (86%) react negatively when advertisers follow their behavior across websites.

Existing research on private information in app markets is based on experimental and survey data. Experimental studies found conflicting valuations for concealing private information. Depending on their framing, the valuations ranged from zero to very large numbers (Carrascal et al., 2013; Beresford et al., 2012; Gros#klags and Acquisti, 2007; Racherla et al., 2011; Tsai et al., 2011). Gros#klags and Acquisti (2007) contrast consumers' willingness to pay to protect their personal information with their willingness to accept for giving it away. The willingness to pay is much lower. Savage and Waldman (2014) use survey-based methods to study consumers' self-reported willingness to accept for giving away personal information on mobile apps. They find an average valuation of $4 for information that is typically shared with app developers. Egelman

---

[4]For a survey of theoretical and empirical studies on the economics of privacy, see Acquisti et al. (forthcoming).

[5]In contrast to empirical evidence, several theoretical models have analyzed the role of private information in online markets. In such models firms use their knowledge about an agent's preferences 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 (Hann et al., 2008; Johnson, 2013b). 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 disadvantageous for consumers. 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 low. Reduced privacy allows to provide many valuable services "for free" and can create benefits for users. Casadesus-Masanell and Hervas-Drane (2015) studied this ambiguity in a model where suppliers compete on two dimensions (price and privacy) in a two-sided market, as it was pioneered by Armstrong (2006) or Rochet and Tirole (2003).

et al. (2013) find that the choice architecture of the platform affects smartphone users' willingness to pay premiums to limit their personal information exposure.[6] High personal information exposure through a smartphone can lead to identity theft and might also have large negative externalities. When users of firm devices grant too many rights to a smartphone app, they might compromise information security. Their usage could open an opportunity for standardized attack patterns like the "Choice and Chance" attack described by Ransbotham and Mitra (2009). Users' adoption decisions can even affect the privacy of uninvolved friends (Pu and Grossklags, 2016).

Several papers studied the general functioning of app markets. Yin et al. (2014) and Davis et al. (2014) investigated innovation in app markets. Carare (2012) studies the impact of bestseller ranks on app demand. Askalidis (2015) and Chaudhari (2015) examine the impact of large scale promotions on the sales of apps. Most closely related to our work is research by Ghose and Han (2014) who estimate the demand for selected apps. They compare 300 top-rated apps on the Android and Apple platforms. Our study differs in three important ways from their pioneering work. Most importantly, we add a new focus by analyzing the role of privacy in such markets for both demand and supply. We also increase the scope of the analysis ($N = 300,000$) and observe real downloads, rather than sales ranks. Finally, Casadesus-Masanell and Hervas-Drane (2015) analyzed privacy as a "second currency" in a model where suppliers compete in both price and privacy.

The core contribution of our paper lies in highlighting the money-for-privacy trade-off *on both sides* of the market for mobile apps, and analyzing its precise nature. We provide the first evidence of this trade-off based on large and detailed data that covers market-transactions for all apps in the Android Market as of summer 2012. Our data set is unique, because we observe in detail which permissions an app requests before it can be installed. Using this information we shed light on app developers' willingness to trade forgone revenue for personal user information, and on how users' installations are related to privacy-sensitive permissions. We are also the first to analyze apps 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. Finally, we provide insight on long-run outcomes and confirm that important mitigating factors of privacy concerns on the web also matter for mobile apps. Some of these factors are outside reputation, the existence of privacy policies or an app's category (cf. Acquisti et al., 2015). On a side note, we would like to stress the importance of this online market, which has *transformed*

---

[6]More technical studies investigated 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 the potential intrusiveness apps (e.g. Chia et al., 2012; Fahl et al., 2012). Sutanto et al. (2013) suggest storing user information locally (on the device) to overcome the personalization vs. privacy paradox, and document a positive effect on demand.

information exchanges in less than 8 years. So, while the app market offers previously unknown potentials for collecting larger amounts and previously unavailable types of private information about consumers, the market is interesting in its own right.

3 The App Market

In 2007, Apple Inc. introduced the iPhone. The device triggered a radical transformation of mobile communication in which screen-only smartphones replaced their predecessors. One of the main competitive advantages of the iPhone and its successors were large app ecosystems. Apps allow users to tailor their devices to their needs by enabling multiple uses besides traditional phone calls and text-based applications. The following year, 2008, saw the release of the first phone using Google's Android Operating System (Android OS). Although users adopted initial versions slowly, Android gained popularity in 2010, and now dominates the market (in most countries). According to the International Data Corporation (IDC, 2015) the Android OS reached a market share of around 75% in 2012. In 2015, the revenue of the mobile app store was near $40 billion and is expected to reach $100 billion in 2020 (App 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. It serves as a distribution channel for apps, books, movies, music and newspapers. In 2012, the Google Play Store featured approximately 400,000 apps, and the number increased to 1.5 million in 2015. Google distinguishes thirty categories of apps, which we sorted into seven overarching meta-categories, such as Education, Entertainment, Games, Tools & Personalization, Lifestyle, Health and Business.[7] In 2012, the largest meta-category was Tools & Personalization (69,372 apps) which contained e.g. weather and transportation apps. The smallest categories were Health and Business related apps (8,255 and 11,686 apps respectively).

A central feature of the Android app ecosystem is its permission system. This system is specific to the Android OS and provides the setting in which the money-for-privacy trade-off can be meaningfully studied. First, developers can choose among standardized blocks of information, so-called permissions, where some enable access to a user's location, communication, browsing behavior etc. Apps must declare *before* installation which permissions they use and must request the permission from the users. More precisely, the system provides a list of permission names alongside a short explanation for each permission. Users must accept this list and explicitly acknowledge that they are granting these permissions to proceed with the installation. Alternatively, they can cancel the installation if they feel uncomfortable about the set of permissions

---

[7]Table 7 shows how we classified the categories into the seven meta-categories.

requested. Note that such explicit consumer consent to the set of permissions does not exist in Apple's iOS, where this information remains implicit before installation. 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.[8] This large number illustrates the quantity and diversity of information app developers can potentially collect about app users. Figure 5 illustrates the way the permissions were displayed in the Android Market in 2012. Since then, Google introduced several small modifications to how permissions are displayed to the user. Before 2014, the list of permissions provided permission names next to short explanations of the permissions. Since 2014 the system shows only the names of aggreagated permission *groups* (but users can open a more detailed dialogue for each permission group). Still, users must approve of the permission list before proceeding with the installation process.[9]

Finally, developers can monetize their apps via four important channels. According to App-Brain (2016), around 20 percent of the apps are paid apps, whereas the remaining apps are for free.[10] Alternative revenue channels are in-app advertisment, in-app purchases and data trade. The importance of these alternative revenue channels was relatively stable since 2012 except for in-app purchases, which were introduced shortly before our period of observation.[11] In 2012, when we collected our data, the "freemium" model based on in-app purchases hardly existed. Since then the market has seen a marked increase of this model, where users may install the apps for free, but must pay a fee to unlock important functions. The two other channels, in-app advertisement and data trade were already common. Data trading is deemed the more privacy-sensitive way of creating revenue from an app, whereas in-app advertisement is deemed more acceptable by many users.

4 Data and Descriptive Evidence

We first describe our data collection. The descriptive analysis of the second subsection highlights three stylized facts about the money-for-privacy trade-off in the market for mobile applications.

---

[8]Today the count stands at 137 permissions (although the precise contents of some permissions changed; see `http://developer.android.com/reference/android/Manifest.permission.html`).

[9]Very recently, Google allows users to withdraw individual permissions from an app after the installation. However, this affects only the most recent version of the Android OS (Version 6.0, named "Marshmallow"). The resulting effects cannot be evaluated in this paper since yet only very customers use this version of the OS.

[10]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`).

[11]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`).

4.1 Data Collection and Variables

We extracted all publicly available information on as many apps as we could find on the English Android Market website in 2012 and 2014 (later "Google Play Store"). 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 three types of information: a demand measure, a price measure and a measure of apps' 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. It is available 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 had to be approximated, which was achieved using an app's sales rank, but not its real downloads (see e.g. Chevalier and Mayzlin, 2006; Garg and Telang, 2013; Ghose and Han, 2014). In addition we use the number of ratings of an app, which is available as a continuous measure, to improve our baseline demand measure. We exploit the continuous number of ratings to predict a continuous number of installations per app, which we then use in our panel analysis. The second main outcome variable is the price of apps, for which we have precise information (in Euros) for each app.

Identifying the Privacy-Sensitiveness of Apps: To measure apps' ability to collect private information, we take advantage of the fact that, as described before, Google provides precise insights into the permissions an app uses. This feature allows us to understand in a detailed way which functions an app can perform, including functions which allow an app to collect private information about the app user.

Among the 136 permissions available to app developers, some can be considered innocuous with respect to the privacy of the user, while others grant an app access to sensitive information. To identify such privacy-sensitive permissions, 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) a classification based

7

on Google's assessment, and (4) one derived by hiring 400 classifiers at Amazon Mechanical Turk.

Our baseline definition of privacy-sensitive permissions follows Sarma et al. (2012) who analyze the benefits and risks of Android app permissions and classify them according to different risk types. 26 permissions are classified as critical, and among these 13 are considered as being a risk to privacy.[12] 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-sensitive 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-sensitive permissions per app.

Three Alternative Classifications of Privacy-Sensitive Permissions: We test the robustness of our results by employing three alternative definitions of privacy-sensitiveness. The first one modifies our baseline definition by classifying only those permissions within an app category as privacy-sensitive, which are both (a) defined as privacy-sensitive by Sarma et al. (2012) and (b) used below-average by paid apps within an app category. The idea is that permissions used rarely by paid apps are atypical for the app category, and that the respective permission quite likely is redundant for the apps' functionality. Instead, it is more likely that it is used 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' ($D_{GoogleMalicious}$). For 36 of the 136 available permissions, Google's official permission description includes a note that the respective permission might be 'potentially malicious', i.e. that it might harm the user of this app. Seven of these potentially malicious permissions are also privacy-sensitive according to Sarma et al. (2012), while the remaining 29 are not. We define the former group as potentially malicious and privacy-sensitive permissions (cf. Table 8) and summarize it in $D_{MaliciousPrivacy}$.

For a third and independent robustness check we use a categorization that we obtained by hiring more than 400 "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

---

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

unlikely to raise concerns, we classified the permission as unproblematic.

Table 8 summarizes all classifications applied and describes each privacy-sensitive permission. In addition, it shows how we grouped the privacy-sensitive 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 its other apps, top developer status, etc.), the app's description (length, number of screenshots, video). Also, we use the section "users who viewed this also viewed" to 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.

4.2 Descriptive Evidence

Before providing descriptive evidence on the three major stylized facts in this paper, we discuss the main variables in our three main data sets.

4.2.1 Three Data Sets

Table 1 provides an overview of the most relevant variables and describes our most important data sets. The descriptive statistics in the table are shown in groups of two columns. In each group, the left column shows averages for free apps and the right column shows those for paid apps.

The first set of columns (col. 1-2) shows the permission usage in the entire cross-section of apps consisting of 233,811 observations.[13] Free apps are installed more often and have a lower average rating. Crucially, an average free app is more likely to use potentially privacy-sensitive permissions and will require more such permissions than an average payable app. For example, free apps use on average 3.06 privacy-sensitive permissions, whereas paid apps use only 1.33 such permissions. Similarly, 25% of the free apps have at least one category-specific privacy-sensitive permission (not usually used in this category of apps), while only 7% of the paid apps used such permissions. These statistics do not change when comparing the presence of permissions

---

[13]The discrepancy between those 233,811 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.

Table 1: Summary Statistics of Three Data Sets

	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
$D_{MTurkSP}$	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	148239	85572	496	496	52380	9539

Notes: The table provides an overview of the most important variables and shows the corresponding descriptive statistics for the three main data sets 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. Columns 1-2 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. Columns 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 predicted number of installations for the cross-section and for the pairs are computed using a slightly reduced sample where observations with no ratings were dropped.

that Google flags as being "potentially malicious" and when applying the classification that we obtained from hiring more than 400 participants on Amazon Mechanical Turk. A third of the free apps use "potentially malicious" permissions (only 11% of the paid apps). Similarly, 19% of the free apps use a very problematic permission, as classified by our hired participants, while only 7% of paid apps do so.

The second group of columns (col. 3-4) shows the descriptive statistics for our set of most rigorously matched app pairs.[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]

---

[14]These pairs were identified using a sword 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 has more permissions but less functionality. Pairs are also quite useful to the

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. Apps in the matched sample 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 whereas the price is lower than that of the average paid app. For the usage of permissions we find very similar patterns as in the full cross-section.

Lastly, in colums 5-6 we provide summary statistics for our panel data set of apps that change permissions over time. Using panel data will allow us to deal with unobserved app-specific heterogeneity. Especially, unobserved quality is expected to matter in our context, because certain features naturally require permission to access data on 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 allows us to control for such unobserved factors in exploring the relationship between downloads and permissions. 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 entire 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 reveals a consistent pattern: No matter how we look at the data, free apps always use both more permissions, and more privacy-sensitive permissions than paid apps. This pattern is prevalent even when matching the pairs of apps that come from the same developer and have the same functionality.

4.2.2 Three Stylized Facts

In this subsection we present three stylized facts. First, we show that some apps use permissions which are not necessary for their functionality. We conjecture that these redundant permissions are used for monetization purposes. Second, we highlight that privacy-sensitive permissions are more likely to appear in free apps than in paid apps. This indicates a negative relationship between price and permission use. Third, the number of installations is negatively related to privacy-sensitive permissions, indicating that demand is negatively related to such permissions.

1. Redundant Permissions: We use both the full sample of app pairs and the sample of selected app pairs to study whether apps request permissions without needing them for their

---

researcher, because we can find permissions that are certainly redundant; the paid version is the proof that the same app can run without the respective permission. 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 less functionality.

Figure 1: Difference in Privacy-Sensitive Permissions between Free and Paid Twin-Apps

(a) Full Pairs Sample (N = 9582)

(b) Selected Pairs Sample (N = 496)
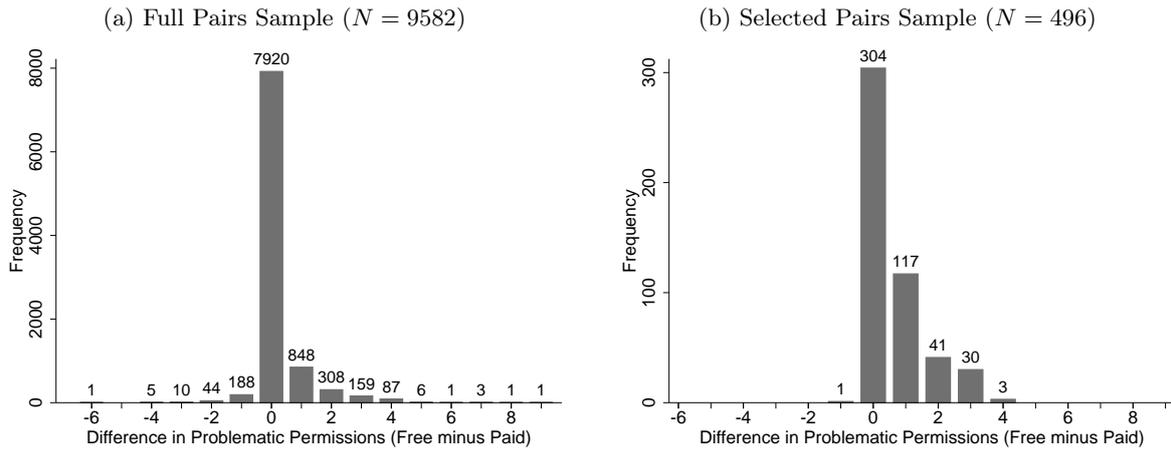
Notes: Both figures compare free and paid versions of an app-pair. On the x-axis we show the differences in the number of privacy-sensitive permissions when comparing the pair's free and paid version (= number of privacy-sensitive permissions of the free app - the number of privacy-sensitive permissions of the paid twin). The y-axis shows the frequency, with which each difference occurs. The left figure shows those frequencies for the full pairs sample, whereas the right figure shows them for the selected sample of app pairs, where we ensure that the functionality between apps does not differ. Most free apps request the same permissions. However, for both samples we clearly see that, if there are any differences, then the free app is more likely to request the additional permission.

functionality. In both data sets, the paid version serves as a technological reference, because the paid version can safely be assumed to provide the same or even *more* 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 compares the number of privacy-sensitive permissions in a pair's free and paid version. Panel (a) shows how frequently we observed a given difference in permissions for the full pairs sample. The frequency 308 for the value +2 means, that we find 308 pairs where the free version used two permissions more than the paid version. Panel (b) shows the same comparison for the selected sample of app pairs, where the functionality does not differ between apps. Across both samples we find more pairs where the free version requires more privacy-sensitive permissions than the paid one. Especially in panel (b) the paid version hardly ever uses more permissions than its free counterpart. We conclude that several of the free versions use redundant, privacy-sensitive permissions, which are not necessary for their functionality.

2. 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. All privacy-sensitive permissions are more common in free apps. Developers choose between offering their apps either for free and with additional privacy-sensitive permissions, or for pay and with fewer privacy-sensitive permissions.

Figure 2: Frequency of Critical Permissions - Free & Paid
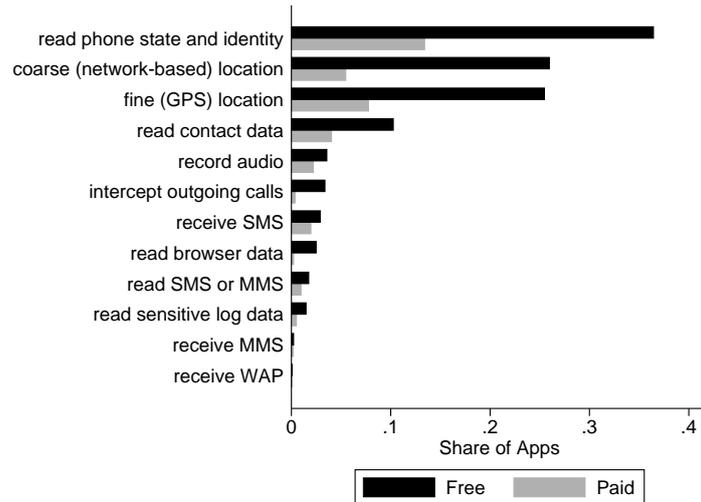
Notes: This diagram shows how often privacy-sensitive permissions are used, both by free and paid apps. Free apps use such permissions much more frequently than paid apps.

3. Demand and Permissions: Figure 3 illustrates our third stylized fact. Users install apps with privacy-sensitive permissions less often than apps which do not use such permissions. The figure displays coefficient estimates of the twelve privacy-sensitive permissions in our data. These estimates are obtained by running simple, descriptive regressions for each of the permissions when controlling for the total number of permissions in the app. The dependent variable is the log-normalized number of installations and the explanatory variable of interest is the respective permission. The coefficients are negative and significant for most of the privacy-sensitive permissions. Apps which use such permissions are installed less often than apps which do not use them.

Figure 3: Number of Installations and Permissions
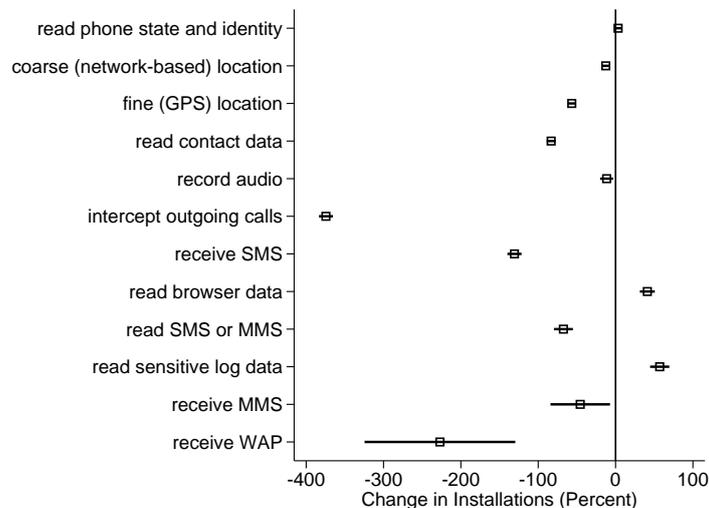
Notes: This figure displays the coefficient estimates of privacy-sensitive permissions. They are obtained with a simple estimation where the dependent variable is the umber of installations in logs and the explanatory variables are the total number of permissions and the privacy-sensitive permission under study. The black lines around the squares indicate 95 percent confidence intervals. The figure suggests that the use of most privacy-sensitive permissions is associated with fewer installations, compared to apps which do not use the permission.

13

The goal of the subsequent sections will be providing rigorous evidence on the three stylized facts and examining their robustness when we control for potential omitted factors.

5 Empirical Framework

This section discusses our empirical framework. The first subsection presents an overview of our estimation approach for the supply side. Subsequently, we discuss estimation of the demand side.

5.1 Supply Side Estimation

We run two types of correlational supply side regressions to provide insights on the role of private information for app developers. First, we examine the choice of the business model. Specifically, we analyze how the developer's decision to offer an app for free or for pay is related to the app's use of privacy-sensitive permissions. Our estimation equation is:

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

where D_i^{Price} is a dummy equal to zero if the app is for free and ε_i is the error term. The dummy D_i indicates whether an app uses privacy-sensitive permissions. If β is estimated negative, apps that use privacy-sensitive permission are $\beta \times 100$ percent less likely to be for pay. Alternatively, a negative β indicates that free apps are more likely to come with a privacy-sensitive permission. The control variables X_i are the same as presented in Table 1: the average rating (in logs), 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 log number of apps of its developer, the average number of installations of the developer's other apps, the minimum and the maximum compatible Android version as well as information about the app's competitors' characteristics.

In a second specification we restrict our sample to paid apps and study how price levels are related to the use of privacy-sensitive permissions. The estimation equation is:

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

where the dependent variable is the log-standardized price of an app. A negative coefficient β indicates that the price level is lower if an app uses any privacy-sensitive permissions. We stress that the regressions in the cross-section do not account for unobserved heterogeneity and only represent conditional correlations.

Panel Data Analysis: To control for unobserved heterogeneity we use a panel specification. We put together a monthly panel covering our observation period of five months. Analyzing the

within-app variation focuses on changes in price or the app's business model and allows us to control for unobserved heterogeneity. However, this approach is limited by the small number of apps that changed price or switched business model during our sample period. For the business model we estimate the following fixed-effects model:

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

where α_i captures unobserved heterogeneity. We estimate the analogous relationship for the price of apps (provided they request payment).

5.2 Demand Side

Our baseline specification to analyze the relationship between app demand and permissions is based on the cross-section sample. We model demand as a function of its permissions, its price P_i and other observable characteristics. 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. Again β is the coefficient of interest, indicating that the use of a certain permission comes with a $\beta \times 100$ percent change in demand. Prices are in logs, so that γ can be interpreted as own-price elasticity.[16] We estimate both a simple OLS and a 2SLS model which accounts for the potential endogeneity of prices, and then proceed to use panel methodology and matching.

Endogeneity of Prices and Permissions: Both, permissions and price are strategic choices of the developer. Hence, the estimated demand coefficients might suffer from endogeneity bias. For monetary prices endogeneity is well understood and usually leads to an upward bias of the estimated price coefficients (Wooldridge, 2010). We use standard instruments for the monetary price such as cost shifters (code length) and the average price of the closest substitutes (Berry et al., 1995; Hausman, 1996; Nevo, 2000).[17] Using such 2SLS techniques, we test whether our coefficients of interest (privacy-sensitive permissions) are affected by the potential endogeneity in prices. For free apps the price is constant and equal to zero and thus instrumentation is not required. Let us now turn to the potential endogeneity of permissions. If developers charge a higher "permission-price" for unobserved high quality, this could introduce a positive correlation

---

[16] A one percent price change comes with a γ percent change in expected demand. We also included D_i^{Price} to compute γ only for paid apps.

[17] (1) Given product quality, code length does not generate utility for the user, but reflects the production cost. (2) We exploit Google's links to other apps that were frequently installed by users who looked at a given app. We did not implement the full set of BLP instruments, because we merely check if instrumentation of price affects the main coefficients. (3) "Hausman instruments" could be used in the panel, but variation in prices is small.

between permissions and the error term. The resulting estimation bias would be in the same upward direction as for the monetary price. We attempt to reduce the effect of unobserved quality using panel estimation techniques and the matched dataset.[18] Any remaining correlation would lead to an upward bias of the regression coefficients, and a truly negative coefficient would be biased towards zero. Without an instrument, we can only provide a lower-bound estimate of the effect of permissions on demand.

Panel: We apply two main strategies to both address the concern of unobserved heterogeneity and to validate our results on alternative data structures. In our first approach we include an app fixed effect (α_i) to estimate our model based on within-app 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. In particular, 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:

$$\Delta Demand_{it+1} = \alpha_i + \beta D_{it} + \gamma P_{it} + \theta X_{it} + \varepsilon_{it}. \tag{5}$$

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

To implement the panel approach we approximate demand with finer measures. This is necessary, because Google's demand measure is too rough to infer monthly downloads which leads to insufficient within-variation over time. We use three alternative approaches to approximate demand. First, we use the predicted number of installations (using the number of ratings as predictor). Alternatively we approximate demand by the number of ratings directly. Second, we run a specification with a developer fixed effect. Third, we restrict our sample to apps which never changed the length of description over the five months we observe. Such apps were more likely to offer 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 price and the number of permissions. App developers often offer two versions of the same app, where one is for free and the other one can be installed after paying a fee. These app-pairs shed light on the money-for-privacy trade-off, as it is perceived by

---

[18]However, if permissions are introduced at a later point, even a panel approach could fail. In our favor, the endogeneity problem is somewhat mediated by the fact that functionality and offering twin-pairs of apps are strategic choices. Developers must make these choices and must introduce the implied permissions before observing demand.

developers. The costly version typically offers some advantage over the free version: It may offer additional functionality, contain less advertisement, and/or be associated with fewer (privacy-sensitive) permissions. Importantly, the paid version serves as technological reference, because any permission that it does not require is not necessary for the functionality of the app.[19]

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 the two apps within a pair are identical in functionality. For that, we need to identify app pairs that have no discernible difference in functionality, and only differ in permissions and price. We manually identify app pairs that state no difference in their app descriptions, or, as the only difference, state that the free version uses advertisements. We then use the differences in permissions and prices within a pair to predict the differences in installations.

6 Results

We first analyze whether cheaper apps use more privacy-sensitive 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 modify the strength of the baseline relationship.

6.1 Money vs. Privacy on the Supply Side

Table 2 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 (col. 1-4) and the price charged given it was positive (col. 5-8). Columns 1-4 analyze the developer's decision to offer an app for money or for free. The dependent variable is a dummy, which is equal to 1 if users have to pay a positive price for downloading the app. Columns 1 and 2 analyze the cross-section, while columns 3 and 4 use the panel of apps which change their business model at least once over the five month sample period. In the cross-section we find that apps which use more privacy-sensitive permissions are between 3.5 and 14.7 percent more likely to be for free. Thus, the correlational results confirm the descriptive evidence showing that "a price comes with fewer privacy-sensitive permissions". This applies to both privacy-sensitive permissions in general and category-specific sensitive permissions. The panel regressions in columns 3 and 4 highlight two patterns. First, less than 1 in 1000 apps ever switches its business model over the period of 5 months. Such a low incidence of switching suggests that the choice of the pricing model is usually a permanent decision, which is revoked only by a small selected sample. Second, if ever such a switch occurs, we find that moving from paid to free coincides with an increase in

---

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

Table 2: Baseline Supply Side Results

	Business Model Choice (D_{Price})				Price Choice of Paid Apps (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 app if for free). In columns 5-8 the dependent variable is the price of an app, given that the developer chose the paid model. In each block, the first two columns analyze the cross-section of all apps, while the second two columns (col. 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 app's pricing policy and our two main measures of an app's use of privacy-sensitive permissions. These measures are category-specific atypical sensitive permissions (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.

the number of permissions and in ad-related permissions. Lastly, we see that internet access and ad-related permissions are more likely in free apps.

In columns 5-8 we analyze the determinants of paid apps' price level. Columns 5 and 6 show cross-section results, while columns 7 and 8 use the panel data. First, note that only around 85,000 apps (36.6%) have a positive price, and of those only 1510 apps (less than 2%) had any variation in price over the five-month period of observation. Similar to the business model, the price of an app is hardly ever adjusted. In the cross-section (col. 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. Such unobserved heterogeneity can be controlled for in panel estimation and indeed we see a negative relationship. While the scope of our panel estimation is limited by the small number of apps which can be included, this negative relationship is in line with the notion that apps might be trading privacy-sensitive 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, because for pay apps request fewer sensitive permissions and fewer category-specific sensitive permissions.

6.2 Demand Side Analysis

Baseline-Specifications and IV: Table 3 shows descriptive regressions analyzing the relationship of app installations and the presence of privacy-sensitive permissions. The dependent variable is the number of installations in logs. Columns 1-5 show cross-section results, with the first column looking at the raw correlation of permissions and installations. Absent any control variables, the coefficient of the privacy-sensitive permission dummy and that of the total number of permissions (including both privacy-sensitive and not privacy-sensitive permissions) have a positive sign which are presumably related to confounding factors. More permissions could be related to greater functionality which then leads to more installations. Once we introduce control variables (col. 2), the positive coefficient of the dummy capturing privacy-sensitive permissions 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 the GPS location for a running app. We identify privacy-sensitive permissions which are atypical for a given category and find that category-specific sensitive permissions are significantly negatively associated with installations. Similarly, when we account for the number of privacy-sensitive permissions as a measure of in-

Table 3: Baseline Demand Side Results

Log. Installations	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 installations and the presence of privacy-sensitive permissions. Columns 1-5 show cross-section results, in column 6 we show 2SLS IV estimation to account for the endogeneity of the price and column 7 analyzes panel data. 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 disaggregates the privacy-sensitive permissions into functionality related types of permissions. Column 6 repeats the specification in column 5, but instruments for price, to account for the likely endogeneity of this variable. Column 7 shows fixed effects panel regressions that use only within-app variation. Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

trusion intensity (col. 4), we also find a significant and negative coefficient.[20] Each additional privacy-sensitive permission would be associated with 10% fewer installations. In column 5 we group privacy-sensitive permissions according to their "kind of functionality". The overall picture confirms the findings in columns 3 and 4: all four permission groups have a negative significant coefficient. Permissions associated with a user's location and communication seem to be the most problematic.

Columns 6 and 7 probe into the robustness of these results when accounting for the endo-

---

[20]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.

geneity of prices or the unobserved heterogeneity of apps. Column 6 shows a 2SLS instrumental variables estimation to account for the endogeneity of price choices. The goal of this specification is to test whether the price 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 average price of the app's competitors. Indeed, the price coefficient becomes more negative in the IV specification, but the coefficients of privacy-sensitive permissions are not heavily affected by instrumenting the app price. Permissions which allow profiling, determining a users' identity or accessing their communication continue to be associated with fewer downloads.

In the last column of the table (col. 7), we run 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, despite the large set of control variables. The results show that privacy-sensitive permissions are associated with fewer downloads. Similarly, the introduction of ads is no longer positive (unlike in the cross-section where ads are associated with greater success). Further reducing the role of unobserved heterogeneity is the main theme of the next section.

Panel Data Analysis and App-Pairs: In Table 4 we address the concern of unobserved heterogeneity in depth. To do so we use two strategies: In columns 1-6 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.[21] Columns 1-2 show the analysis for the full panel of apps which changed their permission usage at least once. In columns 3-4, we further restrict the data to apps which show no change in the (length of the app) description despite changing their permission usage (which would be expected if the app did not change functionality). In columns 5 and 6 we analyze developer-level fixed effects regressions based on 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 adding a developer fixed effect and by restricting the sample to apps of developers which have ten or more apps in 2012. This procedure covers more than 50% of our original cross-section data set.

---

[21]The approximation is based on the categorical variable and the number of reviews. We estimate the nonlinear relationship between installations and reviews (and reviews squared), for free and paid apps separately. We then use these models for a prediction of a continuous installation measure.

Table 4: Panel Demand Side Results

	Panel (ΔPred.Log.Installs.)		Restr.Panel (ΔPred.Log.Installs.)		CS (Dev FE, Log.Installs.)		Pairs (FE, Log.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. In columns 1-4, the dependent variable is the change in predicted log(installations), in columns 5-9, it is the log(installations). Columns 1-4 show the panel that relies on apps that changed permissions at least once between April and October 2012. Downloads are approximated by an estimation procedure based on monthly 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 app pairs 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.

Adding privacy-sensitive permissions comes with a small negative effect in downloads (col. 1 and 3). However, category-specific sensitive permissions (col. 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). When analyzing the regressions with the developer fixed effect (col. 5 and 6) the picture is confirmed for privacy-sensitive permissions. We find a stronger effect than before, but also the standard deviation is increased, because 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 an insignificant coefficient.

In columns 7-9 of Table 4 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.[22] 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. App-pairs are most homogeneous in the sense that these are two versions of the same app, one for free and one for pay. The cost of the increased homogeneity in app-pairs is their scarcity, as we lose many observations, and hence statistical power.

Despite the reduced power in this data set, the dummy for privacy-sensitive permissions is always negative (but also not significant). The total number of permissions is positively correlated with the number of installations for all pairs, but the effect vanishes as we use more and more homogeneous pairs. The results are provided in columns 7-9 of Table 4. Since the paid "twin" is often a premium version of the free version, we expect confounded results when looking at all app pairs we were able to identify. As functionality becomes increasingly similar, however, the coefficients should become less biased. Using all app pairs (col. 7), the point estimate for privacy-sensitive permissions is very small and we find a positive, significant effect for the total number of permissions. If we restrict the pairs sample to those pairs where code size and the description are the same for both app versions (or longer for the free app) (col. 8), this results in a more negative (but still insignificant) effect of privacy-sensitive permissions and a smaller and less significant effect for the total number of permissions. Finally, we use only hand-picked pairs (col. 9) where the paid version does not offer any additional functionality. Here, the positive effect of the number of permissions vanishes, whereas the point estimate for privacy-sensitive permissions remains the same as before (indicating a 10 percent reduction in demand).

Altogether this section documents that users avoid privacy-sensitive permissions in an economically significant way, a result which withstands holding fixed the quality/functionality of an

---

[22]Please refer to the data section for a more detailed description.

app. Across the two approaches in Table 4, panel estimation and matched data, 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.

6.3 Underlying Mechanisms of the Money-for-Privacy Trade-off

We analyze alternative success measures and long-run outcomes, and conclude this section by studying circumstantial factors that modify the negative relationship of permissions and downloads.

Long-Run and Alternative Market Outcomes: Table 5 shows that the results for downloads carry over to alternative and long-run performance measures, such as long term growth and user satisfaction. Specifically, we analyze an app's growth and survival using our long panel covering the period from 2012 to 2014 (col. 1-4), and an app's reputation measured by the average rating and the number of such ratings in 2012 (col. 5-8). Over a 2-year horizon, app survival (col. 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 apps using privacy-sensitive permissions is lower (col. 3 and 4). Columns 5-8 focus on reviews, where the average rating is an indicator of quality and the total number of reviews can be considered an alternative proxy for demand. We find that ratings are lower and fewer, if an app asks for access to privacy-sensitive information. For category-atypical permissions we find an especially strong effect on the number of reviews, which indicates that the negative relationship between permissions and downloads carries over to usage.

Mitigating Factors: In Table 6 we analyze which factors moderate the relationship of privacy-sensitive permissions and installations. We show the results in condensed form here, and the full estimation results in Table 10. 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 (col. 1-3) and contexts where the effect should be larger (col. 4-6). In the lower panel we analyze how the effect varies by category. In these specifications we include a dummy 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 fact that the selected app type might be systematically different. The cross term captures how installations differ with the presence or absence of privacy-sensitive permissions in that group. This analysis serves a

Table 5: 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 (col. 1 and 2) and app growth (conditional on survival; col. 3 and 4) over a 2-year horizon. Columns 5-8 use the cross-section from 2012 and focus on reviews, i.e. the number of reviews (col. 5 and 6) and their average (col. 7 and 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.

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 insights into how and in which contexts privacy concerns matter in the market for mobile applications.

Table 6: 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)

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. It is a condensed summary of Table 10. 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 with 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 permissions that are 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 lower 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.

In columns 1-3 we analyze factors that could reduce users' concerns about their privacy such as the reputation of the organization behind the app. Column 1 studies whether the effect of privacy-sensitive permissions is different for paid apps than for free apps. In paid apps, the effect of sensitive permissions is indeed much smaller, indicating that users trust more when paying

the developer.[23] Column 2 estimates how the relationship of interest differs for apps that have a privacy policy in their app description. 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, etc.). The reputation of a large corporation could 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. For them, 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 at all.

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 are largely in line with the main findings. Column 4 analyzes permissions that are flagged by Google as "potentially malicious". This flag is associated with greater visibility, but possibly also a greater potential intrusiveness. Apps that use such permissions are penalized more strongly than privacy-sensitive permissions not flagged by Google.

Column 5 presents the results from using the alternative permission classification, which we obtained from hiring classifiers on Amazon Mechanical Turk. Based on these classifications, we distinguish unproblematic (reference group), somewhat problematic (SP) and very problematic permissions (VP). Table 6 only shows the coefficients for somewhat problematic (row "No") and very problematic permissions (row "Yes").[27] While apps with somewhat problematic permissions are not installed less frequently than those with only unproblematic permissions, very problematic permissions are associated with fewer installations (-13%). The three permissions which were most frequently classified as problematic were labeled as extremely problematic (EP; cf. Table 10). Their presence is associated with even fewer installations on average (-31%).

Columns 6 and the lower panel present robustness checks which analyze contexts where

---

[23]Alternatively, this may be different selected users who do not mind passing on their credit card data.

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

[25]We use information from Alexa.com on site traffic to generate a dummy which takes the value 1 if we could match an app with a highly ranked website and 0 otherwise.

[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.

[27]We show unproblematic, SP, VP and extremely problematic (EP) in Table 10.

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, because they might have a 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 are less punished for using privacy-sensitive permissions. In the lower panel we differentiated between different categories of apps, because we would expect that the relevance of privacy concerns differs across categories. We distinguish the following categories: Business, Games/Entertainment/Lifestyle, Tools, Education and Health, where Health is the reference category. When looking at the different categories we see that users of Health and Education-related apps seem to avoid privacy-sensitive permissions more strongly than users of Tools, Entertainment or Games.

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 installations, but also to growth, survival, user satisfaction and usage. Second, we find the trade-off to be 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. The reputation of a well established brand creates trust.

7 Discussion

We document an important role of private information as a second currency on both sides of the market for mobile applications: (1) App developers clearly ask for more privacy-sensitive permissions if they offer an app for free than if they offer it for a (higher) price; (2) We observe fewer downloads for apps which ask for more privacy-sensitive permissions (or more sensitive permissions). Other factors being equal, such apps 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 a more privacy-sensitive app or a higher priced less privacy-sensitive app. Consumers choose between these two options (privacy and money) with a certain understanding that there is no free lunch.

We use various subsets of data to obtain our results - a full cross-section, a panel data set, a long panel that tracks apps over a period of two years, and carefully selected pairs of apps. We use three alternative classifications to measure the privacy-sensitiveness of apps (the definition by

Sarma et al., 2012, Google's own classification and one derived through classifiers from Amazon Mechanical Turk). The findings persist for different ways of quantifying intrusiveness (dummy, number of permissions), and we can show the relationship for both supply and demand of apps.

We want to highlight two patterns that emerge from our analysis. First, we find economically significant demand side effects. This is noteworthy for data that predate increased public awareness of data collection practices by the U.S. National Security Administration. At the same time the effects are small compared to previous findings based on surveys (Savage and Waldman, 2014). Our results lie somewhere between the extremes of the privacy valuations found in experimental work and other contexts (Carrascal et al., 2013; Beresford et al., 2012; Grossklags and Acquisti, 2007). A puzzling aspect is that Google's OS makes the information very explicit, and yet the effects on demand are small. On the one hand, the small coefficients could be due to limitations of our data which do not allow to fully isolate the effect of privacy-sensitiveness from functionality. As a result our estimates might underestimate the true reaction of users to privacy-sensitive permissions. Similarly, users might show behavioral biases and overstate their preference for privacy in surveys or experiments. However, Google's warning mechanism seems to trigger much greater avoidance of potentially intrusive mobile apps. Hence, the small effects could also be explained by users who care but do not fully understand.

Alternatively, too much unstructured information about permissions could be like no information at all. This interpretation would align our findings with those of Savage and Waldman (2014), where consumers valued an app's access to private information much more negatively than in our paper, but only *after* they received information about the permissions' technological implications. The large effects of Google's flagging mechanism indicate that simple pieces of summary information might suffice to reduce the gap between earlier findings and ours. One possibility to explore this route further would be introducing an easily interpretable traffic light index to indicate how intrusive an app could potentially be (compare e.g. findings in Tsai et al., 2011). An alternative solution could be to store information locally on the user's device (Sutanto et al., 2013). However, great care is warranted with such changes, because the marked difference in demand for free and paid apps highlights the importance of allowing developers to offer an app for free and monetize the private information 'elsewhere.' This possibility could be a very important driver of both the supply and the adoption of apps. The app market is a potentially quite sensitive two-sided market which has seen an impressive overall performance. Hence, further research must carefully evaluate changes in how users can evaluate the potential intrusiveness of an app before putting them into practice.

The second aspect we want to stress emerges from our analysis of moderating factors. We

confirm users' general negative attitude to privacy-sensitive permissions, but also show that this finding does not always hold. The strength of the relationship depends on the context as well as app and user characteristics, which confirms previous research (Goldfarb and Tucker, 2012; Acquisti et al., 2013). For example, if an app is offered for a positive price, the relationship of privacy-sensitive permissions and downloads is weaker. The negative effects of permissions are also weaker if privacy-sensitive permissions are common in an app's category and if a consumer is likely to be relatively young. However, the negative relationship between demand and permissions can also become stronger. Context and visibility (ease of interpretation) 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 Health or Business apps), intrusive permissions are associated with even fewer downloads. This finding suggests that users are aware of the value of specific types of personal information and share it more cautiously. Ironically, this behavior might hinder the provision and usage of very useful services which require access to personal information.

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. Consumers who care but do not fully understand could create this pattern, because they would rationally prefer a known brand and distrust an unknown app. On the positive side, such behavior allows established firms to improve their products based on consumer data which can increase 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, reinforcing the arguments in Campbell et al. (2015). Such a barrier to entry could lead to reduced 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 robustness of our main results, we do not have the ideal data, which would require access to information on exact download numbers and even more detailed app characteristics. Lacking this information, we have to approximate these variables which always involves compromises, such as a small data set and mostly insignificant coefficients on the most closely matched data set. Moreover, the mitigating factors we analyze do not change over time and hence do not allow for a panel analysis. Hence, these additional results are based only on the cross-section and represent conditional correlations. For example, the weaker negative relationship between permissions and downloads for relatively young consumers should be validated with individual usage data. Such information is needed to decide whether young users install apps too lightheartedly or whether older users show too little trust.

Despite the weaknesses in our data, we see our findings as a first step towards understanding the role of privacy in app markets. Any policy implications that we suggest should be validated with individual level or experimental data. We believe that such a careful evaluation would be a fruitful avenue for further research. 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. This market *transformed* information exchanges in less than eight years and offers previously unknown potentials for collecting private information about consumers. 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 apps in Google's Android Market in 2012 (around 300,000 apps, repeatedly collected in 2012 and 2014). We combined this information with additional data from Alexa.com and from classifiers we hired through Amazon Mechanical Turk. With these data we 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 privacy-sensitive 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 this market: App developers offer either a lower price for a more privacy-sensitive app or a higher priced less privacy-sensitive app. Consumers choose between these two options (privacy and money) and understand that there is no free lunch. Yet, we also find a weakly negative and economically significant relationship between permissions and installations across various specifications. This finding is even more noteworthy, because we collected our data before public awareness to governmental data collection practices was increased by heightened media attention.

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 (such as Business or Health apps). The negative correlation between installations and permissions disappears when the app is associated to reputable web sites or developers. 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 know.

30

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 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. Such a barrier could point to a suboptimal innovative performance of the market for mobile apps especially in "serious" categories like Health or Business. If consumers 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," *Journal of Economic Literature*, forthcoming.

_ , **Laura Brandimarte, and George Loewenstein**, "Privacy and Human Behavior in the Age of Information," *Science*, 2015, *347* (6221), 509–514.

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

App Annie, "App Annie Mobile App Forecast: The Path to $100 Billion," 2016. Available at `https://www.appannie.com/en/landing/forecast`.

AppBrain, "Google Play Stats," 2016. Available at `http://www.appbrain.com/stats/free-and-paid-android-applications`.

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?," Technical Report 2015.

Beresford, Alastair R, Dorothea Kübler, and Sören Preibusch, "Unwillingness to Pay for Privacy: A Field Experiment," *Economics Letters*, 2012, *117* (1), 25–27.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

Campbell, James, Avi Goldfarb, and Catherine Tucker, "Privacy Regulation and Market Structure," *Journal of Economics & Management Strategy*, 2015, *24* (1), 47–73.

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

Carrascal, Juan Pablo, Christopher Riederer, Vijay Erramilli, Mauro Cherubini, and Rodrigo de Oliveira, "Your Browsing Behavior for a Big Mac: Economics of Personal Information Online," in "Proceedings of the 22nd international conference on World Wide Web" International World Wide Web Conferences Steering Committee 2013, pp. 189–200.

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

Chaudhari, Harshal, "The Impact of Zero-Price Promotions on Sales: A Case Study of Amazon Appstore," Technical Report 2015.

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 H, Yusuke Yamamoto, and N Asokan, "Is this App Safe? A Large Scale Study on Application Permissions and Risk Signals," in "in" 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," Technical Report 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 Tucker, "Privacy Regulation and Online Advertising," *Management Science*, 2011, *57* (1), 57–71.

_ **and** _ , "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 "Sixth Workshop on the Economics of Information Securtiy (WEIS)" 2007.

Hann, Il-Horn, Kai-Lung Hui, Sang-Yong T Lee, and Ivan PL Png, "Consumer Privacy and Marketing Avoidance: A Static Model," *Management Science*, 2008, *54* (6), 1094–1103.

Hausman, Jerry A, "Valuation of New Goods Under Perfect and Imperfect Competition," in "The Economics of New Goods," University of Chicago Press, 1996, pp. 207–248.

IDC, "Worldwide Quarterly Mobile Phone Tracker," 2015. Available at `http://www.idc.com/prodserv/smartphone-os-market-share.jsp`.

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

Miller, Amalia R and Catherine Tucker, "Privacy Protection and Technology Diffusion: The Case of Electronic Medical Records," *Management Science*, 2009, *55* (7), 1077–1093.

Nevo, Aviv, "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 2000, *9* (4), 513–548.

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.

Pu, Yu and Jens Grossklags, "Towards a Model on the Factors Influencing Social App Users' Valuation of Interdependent Privacy," *Proceedings on Privacy Enhancing Technologies*, 2016, (2), 61–81.

Racherla, Pradeep, Jeffry S Babb, and Mark J Keith, "Pay-What-You-Want Pricing for Mobile Applications: The Effect of Privacy Assurances and Social Information," in "Conference for Information Systems Applied Research Proceedings," Vol. 4 2011, pp. 1–13.

Rainie, Lee, Sara Kiesler, Ruogu Kang, Mary Madden, Maeve Duggan, Stephanie Brown, and Laura Dabbish, "Anonymity, Privacy, and Security Online," Technical Report 2013.

Ransbotham, Sam and Sabyasachi Mitra, "Choice and Chance: A Conceptual Model of Paths to Information Security Compromise," *Information Systems Research*, 2009, *20* (1), 121–139.

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: Evidence from Smartphone Applications," Technical Report 2014.

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

Sutanto, Juliana, Elia Palme, Chuan-Hoo Tan, and Chee Wei Phang, "Addressing the Personalization-Privacy Paradox: An Empirical Assessment from a Field Experiment on Smartphone Users.," *Mis Quarterly*, 2013, *37* (4), 1141–1164.

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," Technical Report 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.

Turow, Joseph, Jennifer King, Chris Jay Hoofnagle, Amy Bleakley, and Michael Hennessy, "Americans Reject Tailored Advertising and Three Activities that Enable It," Technical Report 2009.

Wathieu, Luc, "Privacy, Exposure and Price Discrimination," Technical Report 2002.

Wooldridge, Jeffrey M, *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press, 2010.

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.

Tables and Graphs

Table 7: Classification of Categories into Seven Meta-Categories.

category	b	pct	meta-category
Personalization	28819	12.3	(1) Tools & Personalization
Entertainment	25179	10.8	(2) Entertainment
Tools	20558	8.8	(1) Tools & Personalization
Books & Reference	15624	6.7	(3) Education
Brain & Puzzle	14687	6.3	(4) Games
Lifestyle	12238	5.2	(5) Lifestyle
Education	11314	4.8	(3) Education
Travel & Local	10419	4.5	(5) Lifestyle
Arcade & Action	8308	3.6	(4) Games
Productivity	8215	3.5	(1) Tools & Personalization
Casual	7963	3.4	(4) Games
Music & Audio	7907	3.4	(2) Entertainment
Sports	7715	3.3	(5) Lifestyle
Business	7239	3.1	(6) Business
Communication	5874	2.5	(1) Tools & Personalization
Health & Fitness	5551	2.4	(7) Health
News & Magazines	5160	2.2	(2) Entertainment
Social	4926	2.1	(2) Entertainment
Finance	4447	1.9	(6) Business
Media & Video	3572	1.5	(2) Entertainment
Photography	2769	1.2	(1) Tools & Personalization
Medical	2704	1.2	(7) Health
Shopping	2525	1.1	(5) Lifestyle
Transportation	2150	0.9	(1) Tools & Personalization
Cards & Casino	2105	0.9	(4) Games
Sports Games	1446	0.6	(4) Games
Comics	1413	0.6	(2) Entertainment
Libraries & Demo	1301	0.6	(3) Education
Weather	987	0.4	(1) Tools & Personalization
Racing	698	0.3	(4) Games
Observations	233813		

Notes: The table shows the frequency distribution of Google's 30 categories, and how they were classified into 7 meta-categories. The categories are sorted by frequency, with the most frequent on top. Tools & Personalization is the largest and Health the smallest meta-category.

Table 8: Permission Group Definitions

Permissions (Group)	Description (provided by Google)	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.

Table 9: Alternative Market Outcomes and Success Measures (Full Table)

	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: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 10: Mitigating Factors (Full Table)

	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 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 (potentially malicious) are penalized more strongly. Column 5 looks at how the coefficient estimates differ for apps that are not suitable for children or young adults. Column 6 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.

Figure 4: App Information in the Android Market 2012

Figure 5: Permission Information in the Android Market 2012

A Online Appendix

A.1 Classification Procedure through Amazon Mechanical Turk

We now describe in more detail how we hired 413 individuals through Amazon Mechanical Turk to obtain an alternative classification of permissions based on the input of these classifiers. We designed an online survey that consisted of a few questions on the classifiers demographic background and then asked them to evaluate permissions. We tested the 12 privacy sensitive permissions alongside with 9 other permissions which are generally considered harmless. We used either permissions which are frequently required for showing ads, and also added permissions which were classified as harmless by Sarma et al. (2012) for validation purposes. To avoid overload, classifiers were not presented with the full list, but only with a random selection of permissions. To avoid framing issues we asked the classifiers 2 questions:

1. How likely is it that you would continue to download an app [...] that requests the following permissions: [Indicate a number from 1 to 10; (10: "without any hesitation", 1. "definitely not")]

2. Would the fact that, everything else equal, an app requests one of the following permissions increase or decrease the likelihood that you install the app.[28] [Three radio buttons "increase"-"same"-"decrease"]

The first question gives a somewhat more precise and nuanced measure for drawing the line between what users are comfortable with. The second question has the advantage of being more straight forward. In the first question, we also iterated the wording to allow verifying if any wording influenced the classification, in the second we used two framings (positive/negative). Finally we asked them to rank the permissions to be able to check the consistency of the classifiers's answers. The entire survey took approximately 10 minutes on average. A copy of the survey can be accessed online.[29]

The data collection took place in early January. We tested the survey with 43 classifiers and, after seeing that the tests were satisfactory, subsequently ran a large collection with 370 individuals.[30] Each classifier could take the survey only once.

Aggregating the answers we ordered the permissions according to how likely they were to incite hesitation. Note that we only care about the relative ordering of permissions. The reported levels of how problematic apps are was not our focus.[31] From the averaged relative ranking of the permissions we then proceeded to order the permissions into extremely problematic (EP), very problematic (VP) and somewhat problematic (SP) ones.[32] These dummies could then be exchanged for the dummies that we obtained from the other two classifications, i.e. from those based on Sarma et al. (2012) or based on Google's classification. To test the robustness of our own classification, we also use a second classification based on question two. Again we generated an ordering of the permissions into extremely problematic (EP_2), very problematic (VP_2) and somewhat problematic (SP_2) ones. Finally we used a third dummy classification with only unproblematic (UP) and problematic (P) permissions. The results we get with those alternative classifications are extremely similar to the results based on the first classification we present (EP, VP, SP), and hence we do not show them for reasons of space.

---

[28]To test framing effects, this question was also framed as 'Are you at ease with installing an app that is able to do the following on your phone? [Three radio buttons "yes"/"no"/"uncertain"]'

[29]at `http://limesurvey.zew.de/limesurvey/index.php/survey/index/sid/784542/newtest/Y/lang/en`

[30]one "production test" and the "full batch" with 290 individuals.

[31]Even though we took several steps to reduce the effects of framing, we do not need the reported levels of privacy concern to be representative. In fact, we believe that the self reported answers outside the context of a real installation are likely to differ from actual choices. Since we were not able to collect experimental data, that is based on real apps and which tracks actual installation behavior, we designed our classification survey to obtain a relative ordering.

[32]For this dummy, we used the answers to question 1. Unproblematic are the permissions that are not mentioned in Sarma, or that received a score above 5. Permissions which received an average score below 4 are somewhat problematic, and permissions with a score below 3 are in the very problematic group (only 4 permissions).

Table A1: Demand and Supply Side Results Using Amazon Mechanical Turk-Dummies

	Supply Side				Demand Side				
	D_{Price}		Log. Price		Log. Installations (in 1000)				
$\#_{TotalPerm.}$	0.004*** (0.000)	0.003*** (0.000)	0.019*** (0.002)	0.015*** (0.002)	0.010*** (0.002)	0.020*** (0.002)	0.019*** (0.002)	0.025*** (0.002)	0.019*** (0.002)
$D_{Internet}$	-0.203*** (0.003)	-0.195*** (0.003)	0.077*** (0.007)	0.082*** (0.007)	-0.014 (0.012)	-0.024** (0.012)	-0.032** (0.012)	-0.038*** (0.013)	-0.019 (0.013)
D_{Ads}	-0.140*** (0.002)	-0.136*** (0.002)	-0.022** (0.009)	-0.022*** (0.009)	0.105*** (0.011)	0.084*** (0.012)	0.095*** (0.011)	0.081*** (0.012)	0.090*** (0.012)
D_{Other}	0.053*** (0.002)	0.059*** (0.002)	0.217*** (0.007)	0.222*** (0.007)	0.039*** (0.011)	0.042*** (0.011)	0.018* (0.011)	0.024** (0.011)	0.036*** (0.011)
D_{Price}					-2.361*** (0.123)	-2.376*** (0.123)	-2.371*** (0.123)	-2.379*** (0.123)	-2.377*** (0.123)
Log. Price					-0.066*** (0.010)	-0.065*** (0.010)	-0.066*** (0.010)	-0.066*** (0.010)	-0.065*** (0.010)
$D_{MTurkSP}$	-0.011*** (0.002)		0.117*** (0.010)		-0.012 (0.012)				0.014 (0.013)
$D_{MTurkVP}$	-0.093*** (0.003)		-0.019 (0.015)			-0.171*** (0.015)		-0.131*** (0.016)	
$D_{MTurkEP}$	-0.114*** (0.005)		-0.069*** (0.026)				-0.369*** (0.028)	-0.310*** (0.028)	
$D_{MTurkSP2}$		-0.040*** (0.002)		0.125*** (0.010)					-0.040*** (0.013)
$D_{MTurkVP2}$		-0.078*** (0.004)		-0.063*** (0.020)					0.044* (0.023)
$D_{MTurkEP2}$		-0.024*** (0.004)		0.104*** (0.022)					-0.186*** (0.024)
Constant	-0.102*** (0.036)	-0.077** (0.036)	-0.030 (0.108)	-0.026 (0.108)	2.986*** (0.213)	2.931*** (0.213)	2.887*** (0.214)	2.858*** (0.214)	2.936*** (0.213)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	233811	233811	85572	85572	233811	233811	233811	233811	233811
Mean of dep. Var.	0.37	0.37	0.27	0.27	-0.66	-0.66	-0.66	-0.66	-0.66
SD of dep. Var.	0.48	0.48	0.77	0.77	3.49	3.49	3.49	3.49	3.49
Adjusted R^2	0.403	0.401	0.271	0.271	0.691	0.691	0.692	0.692	0.691

NOTES: $D_{MTurkSP}$ indicates that the app uses at least one permission classified as somewhat problematic by the classifiers hired through Amazon Mechanical Turk. $D_{MTurkVP}$ indicates very problematic permissions. $D_{MTurkEP}$ indicates extremely problematic permissions. $D_{MTurkSP2}$, $D_{MTurkVP2}$ and $D_{MTurkEP2}$ follow the second classification we used.

Table A1 shows our results when we use the alternative "MTurk sensitivity dummy" in the main specifications. The first four columns show the supply side, while columns 5-9 show the demand side results. Let's first turn to the supply side. In columns 1 and 2 we repeat the specification (1) of Table 2, which shows that using sensitive permissions is associated with a lower likelihood that the app is offered for a fee (or, free apps are more likely to ask for sensitive permissions). When using the new classification we find negative coefficients for any permissions which is flagged by the new dummies. The coefficient is small for somewhat problematic permissions and increasingly strongly negative for very and extremely problematic permissions. In column (2) we use the second classification we derived from the classifiers' work, and get a similar though slightly more blurry picture. In this specification, very problematic permissions (VP_2) are even less likely than the new extremely problematic permissions (EP_2) to be used in for pay apps. This could be driven by one or two problematic permissions that developers of paid apps like to use. Using the two-step classification (not shown) "Unproblematic" permissions are not less likely to occur in apps that are for pay, but problematic permissions are.

Columns 3 and 4 complement the first two columns by asking how price is related to permissions, conditional on the app being for pay. They repeat column (5) of Table A1, but we, again, replaced the indicator based on Sarma et al. (2012) by the indicators based on the work of the classifiers we hired ourselves. Somewhat problematic permissions go together with a higher price, and extremely problematic problematic permissions come with a lower price. As before, this pattern is confirmed for the second classification with the exception that at least some extremely problematic permissions are associated with higher prices. For the two-dummy classification (not shown) we again found no differences with the first classification. Not very problematic permissions come with higher prices, but problematic permissions do not. We want to be sure to clarify that this analysis can only shed light on the raw correlations. A rigorous estimation of causal effects would require at least the use of IV-strategies, but we did not add IV specifications, because the purpose here is to confirm the main findings of the earlier results (and because prices are extremely stable in general). Columns 5-9 replicate our estimations in Table 3. In columns 5, 6 and 7 we estimate the new three coefficients of interest separately, before estimating the coefficients jointly (in column 8). Apps which have one or more somewhat problematic permissions (SP) are not associated with significantly fewer downloads, but apps which command a very problematic permission show -13% installations, whereas extremely problematic permissions show even -31% installations. Like before, this picture is confirmed when we use the second classification from the classifiers. The line between somewhat and very problematic permissions is more fuzzy, when we use this classification, but especially extremely problematic permissions continue to be harshly punished. These results also put the puzzling findings on extremely problematic permissions in columns (2) and (4) into context. If some developers try using extremely problematic permissions in high price apps, this strategy does not necessarily result in more downloads. The two-dummy classification, which we also used but do not show, confirms the findings from columns 5-9.

A.2 Analysis by Category

Tables A2 shows the results for estimating the parameter of interest for each category separately. Before discussing differences in the role of privacy sensitive permissions across app categories we have to point out that the categories differ strongly with respect to installation numbers and prices. For business and health related applications the estimation coefficient for charging a price is only weakly negative, and for educational apps the dummy for a positive price is even positive. This suggests that paid apps are not much less successful than free apps in these markets. However, for games, tools lifestyle and entertainment related apps there is a large penalty for charging a price.

Table A2: Demand Side Results By Category

	Education	Entertainment	Games	Tools	Lifestyle	Health	Business
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\#_{TotalPerm.}$	-0.026***	0.022***	0.027***	0.037***	-0.017***	-0.053***	0.020***
	(0.007)	(0.004)	(0.007)	(0.003)	(0.005)	(0.012)	(0.005)
$D_{Privacy}$	-0.125***	-0.151***	0.016	0.062**	0.234***	-0.053	-0.016
	(0.038)	(0.027)	(0.030)	(0.024)	(0.035)	(0.063)	(0.052)
$D_{Internet}$	0.030	-0.048	0.028	-0.196***	-0.029	0.142**	-0.025
	(0.035)	(0.031)	(0.033)	(0.023)	(0.033)	(0.056)	(0.060)
D_{Ads}	0.486***	0.059**	0.191***	0.136***	-0.059**	0.137***	0.093**
	(0.034)	(0.026)	(0.028)	(0.024)	(0.028)	(0.053)	(0.039)
D_{Other}	0.076**	0.054**	0.168***	-0.076***	0.041	0.048	0.010
	(0.033)	(0.023)	(0.031)	(0.021)	(0.028)	(0.053)	(0.041)
D_{Price}	0.889***	-1.989***	-8.621***	-5.404***	-2.811***	-0.697	-1.113*
	(0.246)	(0.387)	(0.490)	(0.259)	(0.281)	(0.426)	(0.613)
Log. Price	-0.326***	-0.080**	0.426***	0.173***	-0.003	-0.189***	-0.122**
	(0.020)	(0.033)	(0.042)	(0.022)	(0.024)	(0.035)	(0.050)
Constant	-0.417	3.050***	8.688***	6.243***	2.759***	0.475	-0.082
	(0.640)	(0.540)	(0.662)	(0.415)	(0.515)	(0.998)	(0.957)
Category	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	28239	48157	35207	69372	32897	8255	11686
Mean of dep. Var.	-2.01	-0.61	0.46	-0.54	-0.93	-0.67	-0.99
SD of dep. Var.	3.84	3.49	3.29	3.46	3.31	3.14	2.78
Adjusted R^2	0.719	0.682	0.689	0.701	0.667	0.682	0.580

NOTES: This table shows the results for estimating the parameter of interest for each category separately. The dependent variable is log(installations) The columns contain apps from the following categories: Education (Column 1), Entertainment (2), Games (3), Tools (4), Lifestyle (5), Health (6) and Business (7). The explanatory variable of interest is the presence of category specific privacy sensitive permissions. OLS Regressions with heteroscedaticity robust standard errors in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

As in earlier demand specifications, the coefficient of interest in Table A2 accounts for privacy sensitive permissions. Interestingly the negative association between sensitive apps and downloads is strongest for Educational apps and the overall result is also negative for Health Related apps. There is no or little effect of sensitive permissions in Tools, Games and Entertainment and only a very small effect in business apps. We also analyzed category specific sensitive permissions, and they showcase very similar patterns when comparing the categories. In line with the results that we presented in the main body of the paper category-specific sensitive permissions show a stronger negative association with downloads. We find this pattern in all categories except for Business apps, and the difference is strongest for Lifestyle and Games.

Unfortunately, there are multiple possible explanations for the findings in Table A2. Maybe users, who play games and use lifestyle apps are more experienced heavy users of their devices, who are able to notice suspicious permissions. Also the negative coefficients in educational apps and the negative coefficient for any type of permission in health related apps deserve further scrutiny. They might be indicative of more educated users' greater scepticism, but better data are needed to better understand the dynamics in these markets.