Privacy, Data and Non-regulatory Privacy Protection: The Case of Apps For Young Children

March 15, 2023

Abstract

Children are increasingly using mobile apps. Being a vulnerable category of users, US law includes stringent privacy regulation to protect children's personal data, In this market, where security is important to protect children, firms also have incentives to protect children privacy. We investigate whether non-regulatory platform privacy protection and the privacy regulation regime influence the collection of users' data. Mobile apps that opt in to Google's "Designed for Families" program generally comply with US privacy regulations related to children. Our results suggest that a platform can help regulate privacy protection, and we find that self-certification program offered by platform can help reduce data collection from children. Apps that created after the introduction of the self-certification are less likely to request sensitive data especially if they are produced by developers originating from countries with high privacy regime. This result suggests that there are reasons to legislate earlier rather than later when it comes to privacy. Especially, strong privacy regime might lead US developers to be reluctant to produce new apps compared to developers located in country with low privacy regime advocate for an early platform regulation.

Keywords: Platform design, Self-certification, Economics of privacy, Apps for young children.

1 Introduction

This paper investigates the question of how platform self-certification influence data collection in a case where privacy protection undoubtedly matters: Data collection of sensitive information from very young children in game and education app market. While mobile apps can provide learning opportunities, children are not capable of providing informed consent to the data collection practices typically related to mobile apps. Thus, this market is characterized by threats to the security and privacy of this vulnerable audience. While user data can be used to improve app content or to offer personalized ads, how digital platforms and app developers use children's data is unclear. Compared to the collection of data on websites, app data collection is more automated and does not distinguish among users who implicitly agree to data collection when they download the app. This means that data can be collected on very young children. Apps which target very young children tend to be simplistic and have content based primarily on images and sounds which makes it easier to bring them to market and allows developers to easily enter in foreign markets. Reflecting the global app economy, developers of children's apps are located across the world. In addition, the simplicity of these apps means this is a market where apps are low-cost to develop (Ghose and Han, 2014), and many developers from many countries compete in this market.

Firms have incentives to protect consumer privacy as it increases security and trust in the firms Lee *et al.* (2011); Goldfarb and Que (2023) especially when data collected involved sensitive data or vulnerable audience. Google Play Store, the largest app store worldwide, introduced in 2015 a self-certification program called "Designed for Families (DFF)" to help parents identify child-appropriate content. Developers who opt in to the program self-declare that the app complies with Google Play Store's internal DFF policy and the US Children's Online Privacy Protection Act (COPPA) legislation. COPPA protects the privacy of American children under 13 years of age and defines sensitive data in the case of children.[1] This has led the US Federal Trade Commission (FTC) to launch several cases aimed at protecting children's privacy and ensure security. In several cases the FTC has emphasized that the law applies to both national and foreign developers. Children related industry like apps and

[1]COPPA law also regulates ads that target children on the basis of their behavior (behavioral ads).

smart toys involved substantive legal activity due to both privacy and security issues Bleier *et al.* (2020). In 2018, the FTC launched a case against Hong Kong-based VTECH in relation to their Kid Connect app, resulting in a $650,000 fine.[2] VTECH manufactures children's toys and VTech was fined $650,000 because collects data on children as part of its digital toy distribution strategy. In February 2019, the Chinese company which owns the TikTok app, one of the most frequently downloaded apps worldwide, was fined $5.7 million for failing to seek and obtain parental consent for the collection of children's sensitive data.[3] In this case, the FTC again stressed that COPPA legislation applied to any apps that might appeal to children. In April 2019, the FTC fined Google and YouTube $136 million for a COPPA violation, and the company was ordered to pay an additional $34 million to New York state in relation to the same case. The allegations were based on the fact that YouTube advertised companies such as Mattel and Hasbro which target children.[4] In a recent case in 2021, the app Recolor was fined by the FTC as it collects children data with parent's permissions. The companies received complaints from parents and users as children were using the app's social media features such as posting selfies and interacting with other users including adults. [5] This shows that litigation risks in market segment with greater privacy protection, such as children are higher Bleier *et al.* (2020).

We collected weekly data on the apps published in the US market available in Google Play Store over the period July 2017 to January 2021. We collected data on both apps that opted into DFF and those that did not to identify apps that appeal to children, we use search terms such as "preschool" and "toddler." Our dataset includes 27,763 apps published in the US market and 11,338 developers located in 128 countries leading to 1,509,000 observations.[6] COPPA protects the privacy of American children under 13 years of age and defines what is sensitive data in the case of children. We use the COPPA definition of sensitive data to

[2] `https://www.ftc.gov/system/files/documents/cases/vtech_file_stamped_stip_order_1-8-18.pdf`. Last accessed, May 31, 2020.

[3] FTC Cases Proceedings 172-3004. Last accessed, May 31, 2020.

[4] `https://www.ftc.gov/system/files/documents/cases/youtube_complaint.pdf`. Last accessed, May 31, 2020.

[5] `https://www.ftc.gov/system/files/documents/cases/1823184recolorcomplaint.pdf`. Last accessed, March 3, 2023 and `https://medium.com/golden-data/recolor-if-you-cannot-pay-your-coppa-fine-now-the-ftc-will-take-the-money-latter-4259c7b4605e`Last accessed, March 3, 2023.

[6] We deleted apps produced by developers which did not indicate their geographical location since this did not allow us to identify developer's country of origin.

determine whether an app requires sensitive data. This international market is characterized by disparities in privacy regulation regime. To measure the effects of national regulation, we identify developer locations based on the address provided in the app listing on the Google Play Store.

An important regulatory enforcement tool in the context of privacy legislation is industry self-certification, which can affect an industry's competitive structure and ensure competitive advantage(Brill, 2011; Lee *et al.*, 2011; Acquisti *et al.*, 2016). Developers choose whether the app should be included in DFF category or not and no additional monetary costs are associated with opt into the program. Developers who opt in to the DFF declare compliance with COPPA, along with other requirements specified by Google Play Store.

Our findings suggest that the self-certification regime is likely to reduce data collection. We find that 25.83% of apps that self-select into the platform's self-certification program request at least one piece of sensitive data from their child users compared to 49.48% of apps which do not opt in. Thus, children privacy protection may sometimes be more effectively advanced by regulators trying to influence global platform policies towards children, rather than by focusing on changing the regulatory regime within a single country. We also investigate whether the stringency of privacy protection for children in the US may have led developers in the US to be more reluctant to develop apps targeted at the children's market, leading to an opening for international developers to secure market share. We then evaluate whether these results are driven by developer privacy regime, or by underlying developer experience. We find evidence that the relative stringency of privacy protection for children in the US may have led developers in the US to be more reluctant to develop apps targeted at the children's market, leading to an opening for international developers to gain market share. We find positive evidence that platform compliance programs improve child privacy protection, especially among developers from countries with laxer privacy regulations. We also find evidence that apps that opt in has an increase visibility in the platform leading to increased reviews and downloads, which may justify why the developer bears the cost of compliance.

Our work builds upon three streams of academic literature. The first stream of literature is on privacy regulation and security issues. An important regulatory enforcement tool in the

context of privacy legislation is industry self-certification, which can affect the competitive structure (Acquisti *et al.*, 2016; Brill, 2011; Jullien *et al.*, 2020; Gopal *et al.*, 2023) as a firm's privacy protection choice leads to a competition-mitigation strategy Lee *et al.* (2011). Johnson *et al.* (2020) evaluate the loss associated to self-certification initiatives in the ad industry. Regulation can have beneficial effects such as increasing consumers' willingness to share information in more regulated environments. Adjerid *et al.* (2015) show that regulation is associated positively to incentives which have a positive effect on development, adoption, and exchange of health information. Tucker (2014) suggests that giving back some control to the user can increase advertising efficiency. In this context, Miller and Tucker (2017) highlight that regulation which gives the user control over his or her personal data increases adoption of medical technologies while regulation which requires user consent has the opposite effect. Providing information related to privacy issues reduces consumer uncertainty and increases willingness to adopt and use digital products Al-Natour *et al.* (2020). Thus, platform can have an incentive to ensure consumer privacy . We contribute to this strand of literature as we show how strong regulated markets can be enforced through self-regulation program. To our knowledge, there is limited literature on the privacy protection of apps aimed at children, the work of Kesler *et al.* (2017) shows that apps that target the 13+ and 16+ age categories are more intrusive compared to apps targeting the "Everyone" category (which includes children and adults). In computer science, the literature is largely concentrate on popular free mobile apps (Reyes *et al.*, 2018; Liu *et al.*, 2016). They show that the majority of apps do not comply with US child privacy regulation. Our paper extends this analysis by studying how platform policies can protect consumer privacy.

The second stream of literature is on the body of work which demonstrates the role played by platform design on the strategies of app developers. Platform initiative aiming at protection consumers' privacy can increase consumers' security but delaying the compliance to non regulatory initiative can reduce apps' market outcomes (Mayya and Viswanathan, 2022). More generally, platform strategy influences dynamics of competition. By exploiting a change Apple App Store's policy related to its product rating system, (Leyden, 2021) shows that this policy change led to higher-quality products but less frequent product updates. Platform design allows developers to strategically decide when to introduce updates to increase

demand (Comino *et al.*, 2019; Deng *et al.*, 2022; Yin *et al.*, 2014). Ershov (2021) investigates how the design of the Google Play Store changed entry dynamics, and shows that splitting the game category into different subcategories reduces search costs and lowers the quality of new entrants. Our paper extends this analysis by studying how platform policies can support regulation and influence developers' behavior.

Finally, our research contributes to research on children's use of the internet. Internet access has mixed effects on education outcomes (Bulman and Fairlie, 2016; Belo *et al.*, 2013). Empirical evidence shows that internet use in schools affects the level of household internet penetration (Belo *et al.*, 2016). Miyazaki *et al.* (2009) study the importance of self-regulation practices for websites that target children in anticipation of regulatory stringency. We contribute to this work by highlighting the participation of children in the mobile app economy.

Our results are important for regulators because of the importance of protecting children's privacy to ensure their security, and because of some of the intricacies of global competition in the digital space. Children's privacy issues are particularly pressing, as among 8-to 12-year-old children interviewed use mobile devices on average 5.33 in 2021. [7] Our results suggest that policies directed towards improving privacy need to be mindful that in a globally competitive market, it may be more advantageous to encourage platform governance of privacy, rather than focusing on national regulations which may be limited in their global reach.

Our findings suggest that the intensity of data collection is heterogeneous across privacy regimes. In the context of the existing literature, our study makes an important contribution related to estimating the effect of how US children's privacy regulation affects national and foreign developers that commercialize their apps in the US market. The scope and depth of our statistics on children's apps data collection are an improvement on the FTC's initial summary statistics (FTC, 2012a,b). In the mobile apps economy (which is increasingly replacing desktop access to websites), collection of data on very young children may be even more pervasive. Many international developers appear not to comply with any child privacy regulation. As well as providing very comprehensive information on automated data collection practices related to very young children, our empirical analysis provides evidence that can inform future policy. Our empirical approach permits study to not only apps

[7] Common Sense Report published by Rideout *et al.* (2022). Last accessed, March 3, 2023.

dedicated to children but also apps with any appeal to children, of relevance in light of the recent FTC decision under COPPA legislation.[8]

The paper is structured as follows. Section 2 describes the data sources and presents the descriptive statistics. Section 3 presents our empirical strategy and our variables of interest. Section 4 shows the econometric results based on different specifications and provides robustness checks. The conclusion follows.

2 Data

2.1 Market for Children's Apps

Reflecting the global app economy, Google Play Store can be accessed by more than 190 countries. Apps in the Google Play Store are automatically released worldwide with automated translation of the app's description unless the developer specifies otherwise.[9] It is very easy to produce and commercialize apps worldwide for children and especially those under five, since these apps are mainly based on images, sounds, and colors.

2.2 Design for Families Program

The industry has responded to appeals for children's online privacy protection via self-certification initiatives such as the DFF program, which aims to protect young consumers by signaling apps' compliance with COPPA rules. DFF was launched in May 2015. Before Google, the iOS App Store introduced the "Kid category" (Apple's 2013 Keynote) to target children under the age of 13.[10]

Developers choose whether the app should be included in this category or not and there are no additionally monetary costs associated to opt in in the program as there are no additional fees associated with registering for this program. Registration in the Google Play

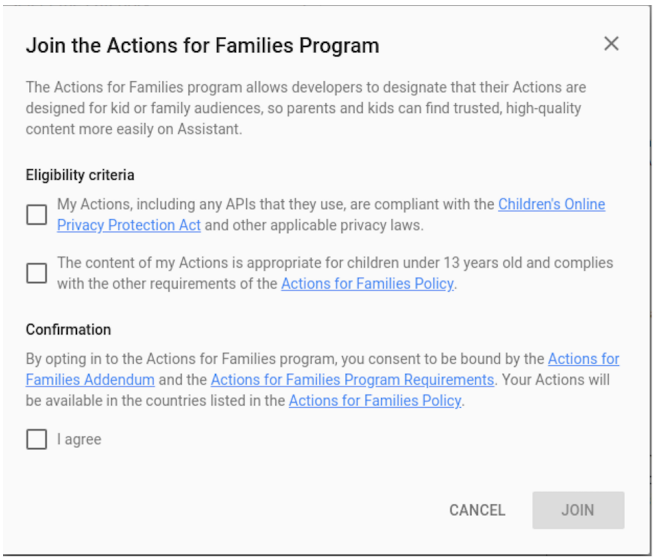
[8] `https://www.ftc.gov/enforcement/cases-proceedings/terms/336`. Last accessed, December 18, 2020.

[9] Certain countries may impose additional requirements on developers to comply with the local regulation.

[10] In June 2019, Apple updated its guidelines for app developers in the kids category and said that they should not include third-party advertising, analytics or links pointing outside the app. `https://developer.apple.com/app-store/review/guidelines/#kids-category`; https://developer.android.com/google-play/guides/families. Last accessed, May 31, 2020.

Store requires the app developer to pay a one-time fee of $25.[11] Developers that include apps in the DFF self-declare that apps comply with platform rules. Google Play Store provides to developers a detailed documentation on app eligibility criteria to belong to this program.[12] Figure 1 shows that consent is based simply on a checkbox indicating agreement for inclusion in DFF.

Figure 1: **Join the Actions for Families Program**



Notes: Eligibility criteria that developers should opt in when joining the DFF.

2.3 Descriptive Statistics

We use data from the Google Play Store. This is the largest worldwide platform that distributes apps for the Android ecosystem. We study children's apps published in the US Google Play Store. We collect weekly data on the full relevant market of children's apps over a three-year period. We follow each app from mid-July 2017 to January 2021, tracking each app starting from its first appearance to the end of the sample period.[13] Our final sample includes 106 weeks as we keep only weeks which contain the full sample of data. We collect data on average every two weeks. The final sample includes 1,509,000 observations with 27,763 apps and 11,338 developers located in 128 countries. This large number of apps

[11] `https://support.google.com/googleplay/answer/143779`. Last accessed, May 31, 2020.

[12] `https://developer.android.com/google-play/guides/families`. Last accessed, July 21, 2020. Certain countries may impose additional requirements on developers to comply with local regulations.

[13] Publicly available data was collected every week via webscraping using the Python programming language.

reflects the fact that producing and commercializing apps for children, especially those under five, is easy due to their reliance on images, sounds, and colors. This is something that has been estimated by Ghose and Han (2014) as part of a broader demand estimation exercise.

Our data collection strategy allows us to collect apps inside the DFF and apps that do not belong to this program using keyword searches aimed at children to capture all children's apps published in the US Google Play Store.[14]

First, we collect the characteristics of apps in the DFF aimed at children aged under 13. It represents 70.6% of our sample.[15] The DFF program includes three broad age categories aimed at children ages 0-5, 6-8 and 9+, with an additional six categories: Action & Adventure, Brain Games, Creativity, Education, Music & Video, and Pretend Play. While the choice of thematic category is optional, developers must choose appropriate age categories.

Second, we construct a benchmark group of apps aimed at children using keyword searches. We identify the list of keywords most frequently associated with children's apps using the Google Adwords keyword planning tool. Table 1 presents the list of these keywords.[16] Google's keyword search algorithm analyzes the app description given by the developer. Google Play search allows users to find relevant and popular apps in the Google Play Store. Algorithmic search is based on title, app description, app icons, images, and screenshots.[17] The search was repeated weekly to identify new benchmark apps. The benchmark group represents 29.4% of the sample.

Apps identified at least once by keyword search in the Google Play Store during the study period are included to our list of apps. This allows us to include broad apps that appeal to children. This aligns with recent COPPA cases, as the FTC declares that general-audience content should comply with COPPA rules if they can potentially appeal to children. Thus, general-audience content are required to comply with COPPA even if it is only particular parts of their websites or apps (including content uploaded by third parties) that are directed

[14] We collect all apps from the search results lists with the maximum scroll-down possible in each page up to the limits of the Google Play Store.

[15] An observation is at app and week level.

[16] In the DFF program, there are 540 apps available in each page and in keyword searches, there are 250 apps available. Apps collected with keywords can overlap with apps inside the DFF. In this case, we consider them as part of the DFF.

[17] App description is the result of developers' strategic behavior. `https://support.google.com/googleplay/android-developer/answer/4448378?hl=en`. Last accessed, November 24, 2020.

at children under age 13.

Table 2 presents descriptive statistics. We collect all publicly available data over time such as type of sensitive data required by apps, number of apps produced by developers, developer addresses, and app characteristics. The Google Play Store provides 21 ranges of downloads for each app from 0 to 5 installs to more than 5 billion installs. We include a set of dummies representing each range (see Table 13 in the Appendix C).

We have an unbalanced panel which allows for entry and exit. New apps appear over time while others become unavailable.

Table 1: **Designed for Family and List of Keywords Used in the Data Collection**

Data Collection Strategy			
DFF Categories	Ages 5 & Under Ages 6-8 Ages 9 & Up Action & Adventure Brain Games Creativity Education Music & Video Pretend Play		
List of Keywords	2 year old	child	preschoolers
	3 year old	children	monitoring
	4 year old	kids	toddler
	5 year old	boy	toddlers
	6 year old	girl	children's
	7 year old	baby	educational
	8 year old	babies	
	9 year old	kindergarten	
	10 year old	kindergartners	
	11 year old	preschool	
	12 year old	kid monitoring	

Notes: The first part of the table presents the list of DFF categories used to collect apps that belong to the program. To each age app category developers can associate any of the categories proposed by the DFF: Action & Adventure, Brain Games, Creativity, Education, Music & Video, and Pretend Play. The second part of the table presents the list of keywords used in the data collection. We use the Google AdWords keyword planning tool which provides keywords most frequently associated with children's apps.

Table 2: **Panel Data Summary Statistics**

	Mean	SD	Min	Max
Dependent Variables				
Sensitive Data	0.586	1.120	0.0	11.0
Prob Sensitive data	0.328	0.469	0.0	1.0
Sharing	0.070	0.254	0.0	1.0
Location Data	0.116	0.321	0.0	1.0
Identity Information	0.244	0.430	0.0	1.0
User Surveillance	0.028	0.166	0.0	1.0
Self-Certification	ref.			
DFF	0.706	0.456	0.0	1.0
App Characteristics				
Contains Ad	0.534	0.499	0.0	1.0
App by developer	18.043	33.714	1.0	248.0
Large # installs	0.079	0.270	0.0	1.0
Privacy Regulation Regime	ref.			
OECD	0.559	0.496	0.0	1.0
US	0.249	0.433	0.0	1.0
Member of the UE	0.302	0.459	0.0	1.0
Recognized by the EU	0.318	0.466	0.0	1.0
Independent authority	0.094	0.292	0.0	1.0
With legislation	0.234	0.423	0.0	1.0
No privacy law	0.052	0.222	0.0	1.0
# Distinct Apps	27,763			
# Distinct Developers	11,338			
Observations	1,509,000			

Notes: This table presents descriptive statistics for the overall sample. Table 13 reports the set of download dummies included into the app characteristics.

3 Empirical Analysis

3.1 Model Specification

We investigate the tradeoffs between promoting competition and protecting children's privacy. We investigate how digital platforms help to enforce legislation requirements. This might differently affect national and foreign developers. This in turn makes the empirical effect of privacy rules ambiguous.

We formalize the key considerations of an app deciding whether or not to request sensitive data. Apps commercialized in the US are produced by US and non-US developers. Each developer faces a binary choice and will decide to enter or not into the DFF. We use variation in privacy regulation worldwide to estimate the effect of different kinds of privacy laws on the types of sensitive data collected. Our empirical work aims to measure the effect of platform policy on protecting children's privacy.

Building on our conceptual framework, we model how self-certification policy are likely to influence the types of sensitive data requested. Our dependent variable, *Sensitive Data*, measures the pieces of sensitive data requested by each app i (i= 1 to N = 27,763) in week t (t= 1 to T=106). We use our panel data to estimate an OLS model with individual app fixed effects, time fixed effects and standard errors clustered on the app level.

We model the intensity of data collection using the following specification:

$$\textbf{Sensitive Data}_{\textbf{it}} = \alpha_0 + D_{it}\omega + \theta_{it} + \zeta_i + \rho_t + \epsilon_{it} \tag{1}$$

Our primary variable of interest is D indicates whether the app i belongs to the DFF program at week t. θ is a vector of other time-varying app characteristics, such as *Contains Ad* and a vector of dummy variable indicating the intensity of download. ζ is the vector of app i fixed effects. Adding the app fixed effects ensures that identification of the coefficient is based on within-app variation over time rather than cross-app variation. The equation also includes time (week) effects (FEs) ρ_t which capture market trends related to privacy over time in our sample. ϵ_{it} is the error term.

3.2 Dependent Variable: Sensitive Data

COPPA regulation defines the list of child-sensitive data collection covered by the law. It includes geolocation details (sufficiently precise to identify street name and city), photos, videos, and audio files that contain children's images or voices, usernames, and persistent identifiers to recognize an app user over time and across different apps.[18] User data can be requested and collected using the permissions system implemented by the Google Play Store. To measure whether children's apps possibly violate COPPA, we identify the Google Play Store permissions and interactive elements (see Appendix A for details) that allow apps to collect these sensitive data on children.

We identify eleven permissions and three interactive elements that require personal data covered by the COPPA regulation. We created the variable *Sensitive Data* which counts the types of sensitive data covered. We identify four broad categories of sensitive data: *Sharing*, *Location Data*,*User Surveillance* and *Identity Information* (see Table 10 in Appendix A to check the permissions and interactive elements required to construct the main dependent variable *Sensitive Data*).

Table 2 presents the descriptive statistics of the main dependent variable. The average number of pieces of sensitive data required by an app is 0.586. We also construct a dummy variable *Prob Sensitive Data* measuring whether the app requests at least one piece of sensitive data; 32.8% of apps belong to this category.

3.3 Self-certification Regime: DFF

A developer's decision to self-certify through the DFF is a strategic choice about customers and competitors (Ershov, 2020). Developers that include apps in the DFF self-declare that their apps comply with the COPPA rules and content is rated "Everyone" or "Everyone 10+" (or equivalent) according to the Entertainment Software Rating Board (ESRB) definition.

[18]The law requires verifiable parental consent for the collection, use, and disclosure of personal information on children aged under 13. This information is not available to the researchers: only developers and users who actually use the app have access to this information. Thus, we are only able to measure the type of permissions required by each app. The complete list of children's personal data is available in FTC rulemaking regulatory reform proceedings (`https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule`). Last accessed, January 8, 2018.

We use the variable *DFF* to identify whether the app belongs to the DFF. Table 3 shows the percentage of apps that collect at least one piece of sensitive data in the group of apps that belongs to the DFF and those that are not in the DFF.

Table 3: **Sensitive Data by DFF**

	Sensitive Data		
	Non-DFF (1)	DFF (2)	Overall (3)
Prob Sensitive Data=1	44.76%	23.79%	29.95%
Mean of dep. var. *Sensitive Data*	1.110	0.368	0.5864

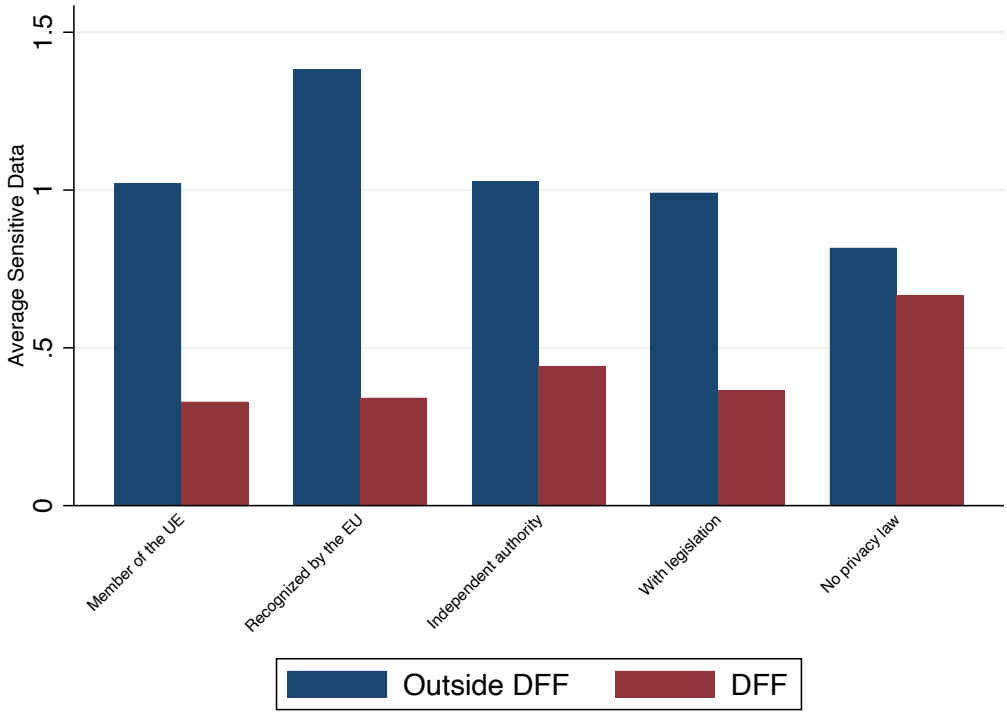
Notes: The table indicates the intensity of sensitive data requested by apps outside and inside DFF.

3.4 Developer's Country and National Privacy Regime

We study children's apps published in the US Google Play Store, but which have been developed worldwide. In our dataset, developers originate from 128 countries. We exploit geographical information disclosed by each developer to identify developer's country. Overall, a plurality of the apps in the US market are produced by US developers (24.95% of the sample). After the US, the largest producers of children's apps are India (with 7.72%), and the United Kingdom (6.31%).

Privacy regulation rules vary across countries, and we exploit this variation to characterize national privacy policies. A developer's privacy strategy might be associated with the home institutional framework. To assess differences in national regulatory frameworks, we augment our data with a vector of the institutional framework measures associated with the developer's address. In the context of privacy regulation, in 1980 the OECD was one of the first international organizations to provide privacy guidelines which were reformed in 2013 (OECD, 2013). Thus, it is reasonable to believe that developers in the OECD have longstanding traditions related to privacy issues. To capture this effect, we create the binary variable *OECD* which identifies developers located in OECD countries. Table 2 presents the intensity of data collection by group of countries. Overall, apps inside the DFF are less likely to collect sensitive data.

Figure 2: **Dff over Privacy Regimes**



4 Results from Panel Data: Sensitive Data Collection from Children

4.1 Main Estimates

Table 4 presents our initial results when we examine how data collection is influenced by self-certification program. We estimate the effect of DFF using: panel FE and time FE, panel FE with time trend, Developer FE and cross-sectional. In each case, the specification includes app characteristics and a vector of dummy variables measuring download intensity.

Column (1) reports the main specification, Equation (1). We include app FE to account for cross-app heterogeneity and week FE for the week the data was scraped. By including FE, we can abstract away from the impact of cross-sectional variation in app characteristics on developer's decision to collect sensitive data. The estimates suggest that apps that opt in to the DFF are less likely to collect child data. If this reflects the ability of platform self-certification initiatives to influence developer behavior, then this program can help with

adherence to local (US) laws. While apps in DFF are not subject to strong enforcement, the platform reminds developers of COPPA legislation requirements (see Figure 1). In the app market, there is fierce competition across all app categories for consumer attention (Bresnahan *et al.*, 2014). The increased visibility in the market for children's apps conferred by DFF certification might compensate for these developers' regulatory compliance costs. Column (2) of Table 4 reports the estimate with app FE and time trend. Column (3) reports the estimate with developer FE and time FE. Column (4) reports the estimate with developer FE and time trend. Column (5) reports a cross-sectional estimates. All estimates show that opt in the DFF reduces sensitive data collection.

Overall, this finding is important from a privacy policy perspective, showing that self-certification is not the only instrument to reduce data collection and might not be sufficient on its own.

Table 4: **OLS Estimates: DFF on Requests for Sensitive Data**

Sensitive Data as Dependent Variable	Main: Apps & Time FE	Apps FE Time trend	Dev FE & Time FE	Dev FE & Time trend	Pooled OLS
	(1)	(2)	(3)	(4)	(5)
DFF	-0.050***	-0.046***	-0.103***	-0.101***	-0.653***
	(0.007)	(0.007)	(0.011)	(0.011)	(0.018)
Constant	0.611***	0.677***	0.599***	0.691***	0.850***
	(0.024)	(0.023)	(0.017)	(0.017)	(0.022)
Apps Characteristics	Yes	Yes	Yes	Yes	Yes
Week Fe	Yes	No	Yes	No	Yes
Dev FE	No	No	Yes	Yes	No
Apps FE	Yes	Yes	No	No	No
Mean Dep.	0.586	0.586	0.586	0.586	0.586
Obs.	1,509,000	1,509,000	1,508,988	1,508,988	1,509,000
Number of groups	27,763	27,763	27,763	27,763	27,763
Cluster	Apps	Apps	Apps	Apps	Apps
Adjusted R2	0.942	0.942	0.806	0.806	0.139

Notes: OLS estimates. *Sensitive Data* is the dependent variable. Robust standard errors are clustered at app level and reported in parentheses. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

4.2 Are Apps in the DFF less Likely to Collect Sensitive Data ?

When we estimate the impact of self-certification, we should compare the decision to opt in to the program considering different control group. In Column (1) of Table 5, we exclude apps

that were always presented in the DFF, we consider only apps that decide to opt in or opt out from the DFF. Column (2) shows the estimates where we exclude the apps that were always presented in the DFF and those that never enter in the DFF in our sample. The coefficients are similar in precision and direction. Column (3) excludes from the estimates developers that have all their apps inside DFF. This permits to test the effect of mix strategy of developers that have some apps in the DFF and other not. Column (4) restricts the estimates on the subsample of apps produced by developers that have at the same time apps inside DFF and apps outside DFF. Column (5) presents the estimates with the look-ahead matching where we match apps using the propensity score matching (Bapna *et al.*, 2018). We compare apps that opt in in the DFF with apps that do not opt in yet but they will join the self-certification program in the future. This approach isolates the analysis from the endogeneity problem, as we only consider apps that will end up opt in and exploit the temporal variation in opt in to identify the impact of the DFF on our variables of interest.

Table 5: **Child Sensitive Data Collection: Effect of the DFF**

Sensitive Data as Dependent Variable	Exclude Apps Always in DFF	Exclude Apps Always in DFF and Never Enter	Exclude Dev with only DFF	Only Dev with DFF and Non-DFF apps at given period	Look-ahead matching
	(1)	(2)	(3)	(4)	(5)
DFF	-0.047***	-0.060***	-0.053***	-0.040***	-0.051***
	(0.008)	(0.008)	(0.007)	(0.014)	(0.010)
Constant	0.891***	0.528***	0.684***	0.517***	0.513***
	(0.028)	(0.029)	(0.018)	(0.032)	(0.005)
Nb Apps by Developer	Yes	Yes	Yes	Yes	Yes
Downloads	Yes	Yes	Yes	Yes	Yes
Contains Ad	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes
Apps FE	Yes	Yes	Yes	Yes	Yes
Obs.	653,438	355,801	1,104,290	337,760	1,079,902
Number of groups	11,268	5,749	19,183	11,024	21,207
Adjusted R2	0.947	0.876	0.943	0.927	0.937

Notes: OLS estimates. *Sensitive Data* is the dependent variable. Robust standard errors clustered at app level are reported in parentheses. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

4.3 Estimates with Alternative Measures of Sensitive Data

We check whether our result holds for different measures of sensitive data. One potential critique is that our main dependent variable includes a broad definition of sensitive data. We check whether a given set of sensitive data is driving our results. Table 6 shows the estimates in each column. We examine the effects of different categories of sensitive data separately.

Column (1) shows the estimates using as dependent variable *Prob Sensitive Data*. Column (2)-Column (5) present estimates with the alternative dependent variable *Sharing,Location Data*, *User Surveillance* and *Identity Information* respectively. Only the estimates of *User Surveillance* is not significant suggesting that DFF is less effective for these types of data. Overall, the results show that apps that belong to DFF might be more careful to share and collect information from this vulnerable audience.

Table 6: **Estimates with Alternative Measures of Sensitive Data**

	Type of permissions				
	(1) Prob Sensitive Data	(2) Sharing	(3) Location Data	(4) User Surveillance	(5) Identity Information
DFF	-0.017***	-0.014***	-0.013***	0.001	-0.010***
	(0.004)	(0.003)	(0.002)	(0.004)	(0.001)
Constant	0.324***	0.078***	0.139***	0.218***	0.039***
	(0.013)	(0.005)	(0.009)	(0.012)	(0.003)
Apps Characteristics	Yes	Yes	Yes	Yes	Yes
Apps FE	Yes	Yes	Yes	Yes	Yes
week	Yes	Yes	Yes	Yes	Yes
Mean Dep.	0.328	0.070	0.116	0.244	0.028
Obs.	1,509,000	1,509,000	1,509,000	1,509,000	1,509,000
Number of groups	27,763	27,763	27,763	27,763	27,763
Cluster	Apps	Apps	Apps	Apps	Apps
Adjusted R2	0.896	0.951	0.909	0.886	0.904

Notes: Linear probability model estimates with app and week fixed effects. Dependent variable as indicated in the table. Robust standard errors are clustered at app level and reported in parentheses. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

4.3.1 Do Experienced Developers that Opt in the DFF Collect less Sensitive Data?

In this section, we check whether a developer's experience as well as the pattern of entry in the market might affect the negative relationship between DFF and sensitive data collection. Given the work of Kummer and Schulte (2019) who find a pattern of developer app experience correlates with requests for more data, it is important to understand how this might influence our results.

We estimate two sets of regressions, dividing the sample according to the year in which

the developer enters the Google Play Store. We consider two distinct groups of developers: those that enter the Google Play Store before the creation of the DFF (May 2015) and those who enter after. We also consider whether each app was created before or after May 2015. Columns (1) and (2) of Table 7 show the estimates of the main equation when we restrict to the sub-sample of apps produced by developers that enter the Google Play Store before May 2015. Column (1) includes only apps created before the creation of the DFF. Column (2) estimates the main equation with the sub-sample of apps created after the creation of the DFF. Note that the DFF program only signals less sensitive data collection relative to the baseline for apps introduced after the DFF program was launched. Column (3) explores what happens when we restrict our sample to sub-samples of apps produced by developers that enter the market after the creation of the DFF (and therefore apps created after May 2015). It shows that the increase of the being in the DFF is negatively associated with sensitive data collection.

Table 7: **Developer Entry Before and After DFF**

Sensitive Data as	Developer Entry Before DFF		Developer Entry After DFF
Dependent Variable	App Created Before DFF	App Created After DFF	App Created After DFF
	(1)	(2)	(3)
DFF	-0.040***	-0.088***	-0.027***
	(0.014)	(0.014)	(0.010)
Constant	0.723***	0.520***	0.600***
	(0.057)	(0.025)	(0.017)
App Characteristics	Yes	Yes	Yes
Week FE	Yes	Yes	Yes
Apps FE	Yes	Yes	Yes
Mean dep. variable	0.703	0.463	0.580
Obs.	488,632	430,580	589,786
Number of groups	7,178	7,154	13,579
Adjusted R2	0.958	0.919	0.936

Notes: OLS with app and week fixed effects. *Sensitive Data* is the dependent variable. Two singleton observations are dropped. Robust standard errors are clustered at app level. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

4.4 Privacy Regulation Regimes

National privacy regime variation across countries is extensive and leads to a wide range of country heterogeneity. We use variation in privacy legislation across countries to estimate the DFF effect within different level of privacy laws. To explore this effect, we split the sample into groups of countries according to stringency of privacy regulation regime.

To account for the heterogeneity of countries in term of privacy regulation, we use the international measure of national privacy regime constructed by the French Privacy Regulation Authority (CNIL).[19] They categorize countries according to their level of compliance with EU privacy legislation (comparable to the US COPPA legislation). Table 12 in the Appendix B presents countries categorized according to their level of compliance with EU privacy legislation. The dummy variable *EU* identifies the developer country as part of the European Economic Area (EEA). The dummy variable *Recognized by the EU* indicates that the country's privacy laws are compatible with EU legislation and thus equally stringent as COPPA. The binary variable *Independent Authority* indicates the existence of an independent authority regulating privacy. The binary variable *With Legislation* indicates that the country has some level of privacy legislation. The dummy variable *No Privacy Law* indicates absence of privacy laws in the developer's country.

The baseline specification for different sub-samples are reported in Table 8. To facilitate the interpretation of the estimates, we report the mean value of the dependent variable *Sensitive data*. Column (1) explores what happens when we restrict our sample to apps produced in the OECD. Apps that opt in the DFF are likely to reduce data requests. The results in column (2) show the regression on the subsample of apps produced in non-OECD member countries. Being in the DFF tends to decrease the pieces of sensitive data collected.

Column (3) displays the results of the sub-sample of apps produced by US developers. Apps commercialized by developers in the US that opt in to DFF are less likely to request sensitive data. The coefficient associated with *DFF* is substantially larger for apps produced in the US compared to other estimates. This provides suggestive evidence that self-certification is more efficient when driven by home regulation.

[19] CNIL, "La protection des données dans le monde". `https://www.cnil.fr/fr/la-protection-des-donnees-dans-le-monde`. Last accessed, January 8, 2018.

Column (4) explores what happens when we restrict our sample to apps produced in EU which has a children's privacy protection regime comparable to COPPA. The estimate shows that only the self-certification regime is likely to affect the pieces of sensitive data requested by European developers.

In the sub-sample of apps produced in countries with a privacy legislation recognized by the EU (Column (5)), we see similar estimates as in column (3). Comparing the point estimates of the DFF coefficients across the columns of Table 8 shows an increase effect of apps that opt in in the DFF for apps produced in the US and countries with legislation recognized by the EU. Column (6) shows the estimates on a sub-sample of apps produced in countries with an independent privacy authority. For this set of apps, DFF does not seem to reduce the pieces of sensitive data requested.

Columns (7) and (8) show respectively that the apps produced by larger developers in countries with privacy legislation (*With Legislation*) and without any privacy legislation (*No privacy regime*) are less likely to collect sensitive data. The estimates show that apps in DFF are less likely to request sensitive data. This suggests that conditional on already having a strong privacy regulatory regime relating to children's data (US and country with legislation recognized by the EU), consumer protections may be more effectively improved by influencing digital platform global policies towards children rather than changing the regulatory regime within a single country.

Table 8: **Intensity of Data Collection and Privacy Regimes**

	OECD vs. Non-OECD		US	Privacy Regime				
	OECD	Non-OECD	US	EU	Rec. EU	Ind. Aut	With leg	No Privacy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DFF	-0.048***	-0.048***	-0.094***	-0.022**	-0.076***	0.009	-0.067***	-0.047*
	(0.009)	(0.010)	(0.017)	(0.010)	(0.014)	(0.017)	(0.016)	(0.026)
Constant	0.573***	0.630***	0.698***	0.481***	0.642***	0.715***	0.614***	0.793***
	(0.017)	(0.022)	(0.028)	(0.026)	(0.023)	(0.046)	(0.028)	(0.056)
App Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Apps FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean dep. variable	0.578	0.597	0.685	0.508	0.647	0.638	0.557	0.706
Obs.	843,735	665,265	376,436	455,393	479,898	142,271	353,359	78,079
Number of groups	14,451	13,312	6,441	7,929	8,187	2,571	7,046	2,030
Adjusted R2	0.954	0.926	0.962	0.942	0.959	0.921	0.920	0.943

Notes: OLS with app and week fixed effects. *Sensitive Data* is the dependent variable. Column (1) shows the estimates within the sub-sample of apps produced in OECD member countries. Column (2) shows the estimates within the sub-sample of apps produced in non-OECD member countries. Column (3) reports the estimates of the sub-sample of apps produced in the US. Column (4) reports the estimates of the sub-sample of apps produced in the EU. Column (5) reports the estimates of the sub-sample of apps produced in countries with a privacy regulation regime recognized by EU. Column (6) shows the estimates within the sub-sample of apps produced in countries with an independent privacy authority. Column (7) shows the estimates of apps produced in countries with a privacy legislation. Column (8) shows the estimates of apps produced in countries with no privacy legislation. Robust standard errors are clustered at app level and reported in parentheses. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

5 What Drive the Decision to Opt In in the DFF ?

Developers that include apps in the DFF self-declare that apps comply with platform rules and the participation to the program has no additional costs. Privacy protection increase users' security but it increases firm compliance costs which might off set the benefit of compliance. In the app market, the platform has a strong market power being the world's largest mobile application platform in term of number of apps commercialized. The market power of platform could increase the incentive of developers to comply to the self-certification. The main benefits to the developer are the reduction of discovery cost and increase of trust from the consumer.[20]

We investigate whether the decision to opt in to the program is associated with increase visibility of a given app in the platform. We use number of reviews and the probability of being a killer app to measure increased app visibility. We use three different dependent variables: *Nbr Reviews*, *Review Growth*, *Killer apps* which measure respectively the number

[20]Apps included in the program will allow parents to find them more easily.

of reviews of a given app, the growth of number of reviews and the probability of being an app with 10% top percentile downloads.

Column (1) of Table 9 shows that being in the DFF is likely to increase the number of reviews, suggesting that developers strategically opt in to DFF to increase their visibility while taking bear the opportunity costs of compliance. This results is also corroborating in Column (3) where we estimate the review growth which is likely to positively correlated with the presence in the DFF. Column (2) estimates the correlation between DFF and the growth of number of review from one week to another. Column (4) estimates the correlation between DFF and being a killer app from one week to another. The correlation is negative and significant suggesting that being in the DFF is likely to decrease the probability to be a killer app. There are two potential mechanisms. Killer apps have already high visibility in the platform (they do not necessarily need to integrate DFF) and they might not want to bear the risk of potential privacy breach.

Table 9: Effect of Opt In on the Number of Reviews and Downloads

	Nbr Reviews	Δ Nbr Reviews	Reviews Growth	Killer Apps
	(1)	(2)	(3)	(4)
DFF	213.884***	2466.770	0.019***	-0.014***
	(76.516)	(1616.810)	(0.001)	(0.005)
Constant	373.285	21293.870*	0.008***	0.137***
	(324.733)	(11036.353)	(0.002)	(0.014)
Contains Ad	Yes	Yes	Yes	Yes
Paid apps	Yes	Yes	Yes	Yes
Freemium	Yes	Yes	Yes	Yes
Obs.	1,510,735	1,482,716	1,320,640	439,715
Number of groups	27,780	27,541	24,822	11,974
Cluster	Apps	Apps	Apps	Apps
Adjusted R2	0.449	0.009	0.660	0.807

Notes: OLS with app and week fixed effects. Dependent variable as indicated in the table. Robust standard errors are clustered at app level and reported in parentheses. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

6 Conclusion

This paper provides empirical evidence that an app produced in the DFF are less likely to collect sensitive data. The question then becomes how best to protect child privacy and ensure their security. Using panel data variation, we show that Google's self-certification program

that allows developers to opt in to self-certify, can help to protect children's privacy.

These results have several implications. First, our results support the view that non regulatory interventions proposed by companies can protect children privacy. Third, our results suggest also that the high standards imposed by regulation can create market distortions by affecting developers in different ways depending on their capacity to comply with the regulation. The platform self-certification regime seems to encourage US developers to comply with COPPA regulation. This finding is aligned with the aim of the platform to encourage compliance with COPPA legislation.

Further research is needed to investigate the extent to which privacy protection is also associated with better content for children. A potential limitation of our findings is that we have no information on the objectives of data collection beyond content improvement and expected users behavior. However, this study provides a first attempt to understand the complexity of the child apps market and how national privacy regulation affects firms' decisions worldwide.

References

Acquisti, A., Taylor, C. and Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*. 54(2), 442–92.

Adjerid, I., Acquisti, A., Telang, R., Padman, R. and Adler-Milstein, J. (2015). The impact of privacy regulation and technology incentives: The case of health information exchanges. *Management Science*. 62(4), 1042–1063.

Al-Natour, S., Cavusoglu, H., Benbasat, I. and Aleem, U. (2020). An empirical investigation of the antecedents and consequences of privacy uncertainty in the context of mobile apps. *Information Systems Research*. 31(4), 1037–1063.

Bapna, R., Ramaprasad, J. and Umyarov, A. (2018). Monetizing Freemium Communities: Does Paying for Premium Increase Social Engagement? *MIS Quarterly*. 42(3), 719–736.

Belo, R., Ferreira, P. and Telang, R. (2013). Broadband in school: Impact on student performance. *Management Science*. 60(2), 265–282.

Belo, R., Ferreira, P. and Telang, R. (2016). Spillovers from Wiring Schools with Broadband: The Critical Role of Children. *Management Science*. 62(12), 3450–3471.

Bleier, A., Goldfarb, A. and Tucker, C. (2020). Consumer privacy and the future of data-based innovation and marketing. *International Journal of Research in Marketing*. 37(3), 466–480.

Bresnahan, T., Davis, J. P. and Yin, P.-L. (2014). Economic value creation in mobile applications. In *The changing frontier: Rethinking science and innovation policy*. (pp. 233–286). University of Chicago Press.

Brill, J. (2011). The intersection of consumer protection and competition in the new world of privacy. *Competition Policy International*. 7(1), 7–23.

Bulman, G. and Fairlie, R. W. (2016). Technology and education: Computers, software, and the internet. In *Handbook of the Economics of Education*. (pp. 239–280). vol. 5. Elsevier.

Comino, S., Manenti, F. M. and Mariuzzo, F. (2019). Updates management in mobile applications: iTunes versus Google Play. *Journal of Economics & Management Strategy*. 28(3), 392–419.

Deng, Y., Lambrecht, A. and Liu, Y. (2022). Spillover effects and freemium strategy in the mobile app market. *Management Science*.

Ershov, D. (2020). *Competing with superstars in the mobile app market*. Working Paper #18-02, NET Institute, USA.

Ershov, D. (2021). *Consumer product discovery costs, entry, quality and congestion in online markets*. Working Paper, Toulouse School of Economics, France.

FTC (2012a). *Mobile apps for kids: current privacy disclosures are disappointing*. Technical report.

FTC (2012b). *Mobile apps for kids: disclosures still not making the grade*. Technical report.

Ghose, P. and Han, S. P. (2014). Estimating demand for mobile applications in the new economy. *Management Science*. 60(6), 1470–1488.

Goldfarb, A. and Que, V. F. (2023). *The Economics of Digital Privacy*. Technical report. National Bureau of Economic Research.

Gopal, R. D., Hidaji, H., Kutlu, S. N., Patterson, R. A. and Yaraghi, N. (2023). Law, Economics, and Privacy: Implications of Government Policies on Website and Third-Party Information Sharing. *Information Systems Research*.

Johnson, G. A., Shriver, S. K. and Du, S. (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science*. 39(1), 33–51.

Jullien, B., Lefouili, Y. and Riordan, M. H. (2020). Privacy protection, security, and consumer retention. *Security, and Consumer Retention (June 1, 2020)*.

Kesler, R., Kummer, M. E. and Schulte, P. (2017). *Mobile applications and access to private data: The supply side of the Android ecosystem*. ZEW - Centre for European Economic Research, Discussion Paper # 17-075.

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

Lee, D.-J., Ahn, J.-H. and Bang, Y. (2011). Managing consumer privacy concerns in personalization: a strategic analysis of privacy protection. *MIS Quarterly*, 423–444.

Leyden, B. T. (2021). *Platform design and innovation incentives: Evidence from the product ratings system on Apple's App Store*. CESifo Working Paper # 9113.

Liu, M., Wang, H., Guo, Y. and Hong, J. (2016). Identifying and analyzing the privacy of apps for kids. In *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications*. February. ACM, 105–110.

Mayya, R. and Viswanathan, S. (2022). Delaying Informed Consent: An Empirical Investigation of Mobile Apps' Upgrade Decisions. *Available at SSRN 3457018*.

Miller, A. R. and Tucker, C. (2017). Privacy protection, personalized medicine, and genetic testing. *Management Science*. 64(10), 4648–4668.

Miyazaki, A. D., Stanaland, A. J. and Lwin, M. O. (2009). Self-regulatory safeguards and the online privacy of preteen children. *Journal of Advertising*. 38(4), 79–91.

OECD (2013). *Privacy expert group report on the review of the 1980 OECD privacy guidelines*. Technical Report 229.

Reyes, I., Wijesekera, P., Reardon, J., On, A. E. B., Razaghpanah, A., Vallina-Rodriguez, N. and Egelman, S. (2018). Won't somebody think of the children? Examining COPPA compliance at scale. *Proceedings on Privacy Enhancing Technologies*, 63–83.

Rideout, V., Peebles, A., Mann, S. and Robb, M. B. (2022). Common Sense census: Media use by tweens and teens 2021.

Tucker, C. (2014). Social networks, personalized advertising and privacy controls. *Journal of Marketing Research*. 51(5).

Yin, P. L., Davis, J. P. and Muzyrya, Y. (2014). Entrepreneurial innovation: Killer apps in the iPhone ecosystem. *American Economic Review*. 104(5), 255–59.

Supplementary Appendix A:

The Dependent Variable

A.1 Descriptive Statistics of Permissions and Interactive Elements Used to Construct Sensitive Data

Sensitive Data is the major dependent variable because it aggregates all types of COPPA-designated categories of sensitive data. It includes four subsets of sensitive data measures: *Sharing*, *Location Data*, *Identity Information* and *User Surveillance*. Table 10 presents the detailed descriptive statistics of each piece of sensitive data used to construct the dependent variable. It also provides detailed statistics by developer location.

The variable *Sharing* takes value 1 if the app requests at least one of the interactive elements allowing apps to share users' personal data with other apps and third parties; this includes *Share Location*, *Share Info* and *Users Interact*. In 2015, the Google Play Store announced the presence of interactive elements to inform consumers on what information the app has access to. The binary variable *Users Interact* measures if the app exchanges sensitive data between users. This feature allows the app to be exposed to unfiltered/uncensored user-generated content including user-to-user communications and media sharing via social media and networks. *Share Info* measures whether the app shares users' personal information with third-parties such as Instagram, Viber and other social networks. *Share Location* equals 1 if the app shares users' locations to other users of social network likes Facebook and Snapchat.[21]

We identify four permissions that request users' location data to construct the binary variable *Location Data*. *ALEC* (Access Location Extra Commands) indicates whether an app collects user's locations based on various device capabilities, and *ANBL* (Approximate Network Based Location) is used to access approximate location derived from network location sources such as cell towers and Wi-Fi. *MLST* (Mock Location Sources for Testing) is used to facilitate developer access to users' locations, and *Precise GPS Location* provides accurate location data.

The binary variable *Identity Information* includes two permissions to identify unique individual identity. The permission *Read Phone Status and Identity* allows developers to identify a smartphone's unique IMEI (International Mobile Equipment Identity) which is considered a persistent unique identifier by COPPA and GDPR (Reyes *et al.*, 2018). The IMEI can be used to recognize a user over time and across different online services,[22] and it could be used to log all kinds of personal data and target the consumer. The IMEI number also permits developers to know which advertising is already seen by a user. A child's voice can be captured via the permissions *Record Audio*.

User surveillance is a binary variable that measures whether at least one permission allows access to user activity and contact information. *Read Your Own Contact Card* allows developers to access users' contact cards and associate users' phone numbers with their names. *RCEPCI* (Read Calendar Events Plus Confidential Information) is used to read information stored on users' phones including those of friends. *Read Your Contacts* indicates whether the app reads users' contacts stored including the frequency with which the user communicates with a given individual. The permission *Read Call Log* allows the app to access data about incoming and outgoing calls. *Read Your Browser History and Bookmarks* gives access to web browser information including internet account information.

[21] See esrb.org. Last accessed, July 21, 2020.

[22] Complying with COPPA: Frequently Asked Questions. Last accessed, September 3, 2020.

Table 10: **List of Permissions and Interactive Elements Used to Construct the Dependent Variable *Sensitive Data***

	Overall (1)	US (2)	EU (3)	OECD (4)	Non-OECD (5)
Sharing	0.081	0.104	0.082	0.090	0.070
Share Location	0.014	0.020	0.013	0.014	0.013
Share Info	0.013	0.013	0.018	0.015	0.011
Users Interact	0.054	0.071	0.051	0.061	0.047
Location data	0.188	0.221	0.155	0.175	0.205
ALEC[a]	0.003	0.004	0.003	0.003	0.003
ANBL[b]	0.096	0.111	0.075	0.089	0.105
MLST[c]	0.001	0.000	0.001	0.000	0.001
Precise GPS Location	0.088	0.105	0.076	0.083	0.095
Identity Information	0.275	0.296	0.238	0.267	0.284
Read Phone Status And Identity	0.199	0.198	0.166	0.180	0.222
Record Audio	0.076	0.097	0.072	0.087	0.062
User Surveillance	0.043	0.065	0.033	0.046	0.038
Read Your Own Contact Card	0.005	0.009	0.002	0.005	0.004
RCEPCI[d]	0.007	0.007	0.005	0.006	0.008
Read Your Contacts	0.022	0.036	0.018	0.025	0.018
Read Call Log	0.005	0.007	0.005	0.006	0.004
Read Your Browser History and Bookmarks	0.004	0.006	0.003	0.004	0.004
Observations	1509000	376436	455393	843735	665265

Notes: This table depicts the summary statistics of the permissions and interactive elements used to construct the main dependent variable *Sensitive Data*. Column (1) presents the overall mean. Column (2) presents the mean for sensitive data requested by apps produced in the US. Column (3) presents the mean for sensitive data requested by apps produced in the EU. Column (4) presents the mean for sensitive data requested by apps produced in the OECD countries. Column (5) presents the mean for sensitive data requested by apps produced in the non-OECD countries.

[a] ALEC: Access Location Extra Commands.
[b] ANBL: Approximate Network Based Location.
[c] MLST: Mock Location Sources for Testing.
[d] RCEPCI: Read Calendar Events Plus Confidential Information.

Supplementary Appendix B:

COPPA Regulation Enforcement and Developer Location

B.1 COPPA Regulations Enforcement

The FTC ensures compliance with COPPA legislation in the US and in other countries. Since COPPA was implemented, the FTC has investigated more than 30 cases. Table 11 presents some recent cases. Some of these cases involve the app developer directly. The FTC imposes strong requirements regarding the type of data that companies can collect, and how they should protect children's personal information.[23]

Table 11: **COPPA Regulations Enforcement**

Firms	Date	Settlement	Country	Mobile Apps
WW International, Inc.	2022	$1,500,000	US	Yes
OpenX Technologies, Inc.	2021	$2,000,000	US	No
Recolor	2021	$3,000,000	US/ Finland	Yes
TikTok	2019	$5,700,000	China	Yes
HyperBeard	2019	$150,000	US	Yes
YouTube[a]	2019	$170,000,000	US	-
Inmobi	2016	$950,000	Singapore	Yes
LAI Systems	2015	$60,000	US	Yes
Retro Dreamer	2015	$300,000	US	Yes
TinyCo, Inc.	2014	$300,000	US	Yes
Path, Inc	2013	$800,000	US	Yes
Artist Arena LLC	2012	$1,000,000	US	No
RockYou, Inc.	2012	$250,000	US	No
Broken Thumbs	2011	$50,000	US	Yes
Playdom, Inc.	2011	$3,000,000	US	No
Skidekids.com	2011	$100,000	US	No
Iconix Brand Group	2009	$250,000	US	No
Imbee.com	2008	$130,000	US	No
Sony Music Song BMG	2008	$1,000,000	US	No
Xanga.com	2006	$1,000,000	US	No
Ms. Fields Famous Brands	2003	$100,000	US	No

Notes: The table illustrates the amount of settlements imposed by FTC under COPPA rules. All cases can be find on the FTC website.

[a] `https://www.ftc.gov/system/files/documents/cases/youtube_complaint.pdf`. Last accessed, May 31, 2020.

B.2 Developer Location

To explore US regulation spillovers to other countries, we retrieve geographical information disclosed by developers of apps available in the Google Play Store. Although the FTC requires that firms collecting or maintaining sensitive data from children should indicate in their online notices or information practices their name, address, telephone and email address,

[23] `https://www.ftc.gov/news-events/blogs/business-blog/2018/10/happy-20th-birthday-coppa`. Last accessed, July 21, 2020.

several developers fail to provide a geographical address.[24]

To retrieve developers' countries, we use different strategies. First, we use Maps APIs to collect the latitudes and longitudes of the given address to identify the country. Second, we used a Python library (Libpostal)[25] to search for a country name in the developer's address. Third, we check the match between the location identified using the Google Maps APIs and the country name identified via Libpostal. Fourth, among the subset of apps without any developer's address, we identify their location using the email extension. Using this procedure, we identify the origin countries of 310 apps. Finally, we manually check for certain addresses. We delete apps produced by developers which did not indicate their geographical location since this did not allow us to identify country of origin. To summarize, 19.22% of the initial sample fall into this category.

[24] `https://www.ecfr.gov/current/title-16/chapter-I/subchapter-C/part-312`. Last accessed March 2, 2022.

[25] `https://github.com/openvenues/pypostal`. Last accessed, February 13, 2020.

Table 12: **Privacy Regime Based on EU Privacy Regulation: List of Countries Presented in Our Sample**

EU	Recognized by EU	Independent Authority	With Legislation	No Privacy Law
Austria	Andorra	Albania	Angola	Afghanistan
Belgium	Argentina	Australia	Armenia	Algeria
Bulgaria	Canada	Bosnia and Herzegovina	Azerbaijan	Bahrain
Croatia	Israel	Colombia	Brazil	Bangladesh
Cyprus	New Zealand	Costa Rica	Chile	Barbados
Czech Republic	Switzerland	Gabon	China	Belarus
Denmark	US[a]	Ghana	India	Bolivia
Estonia	Uruguay	Hong Kong	Indonesia	Cambodia
Finland		Korea, Rep.	Japan	Congo, Rep.
France		Kosovo	Kazakhstan	Cuba
Germany		Macedonia, FYR	Kyrgyz Republic	Dominican Republic
Greece		Mexico	Malaysia	Ecuador
Hungary		Moldova	Montenegro	Egypt, Arab Rep.
Iceland		Morocco	Nepal	El Salvador
Ireland		Senegal	Nicaragua	Ethiopia
Italy		Serbia	Philippines	Guatemala
Latvia		Tunisia	Qatar	Honduras
Lithuania		Ukraine	Russian Federation	Iran, Islamic Rep.
Luxembourg			Seychelles	Iraq
Malta			Singapore	Jamaica
Netherlands			South Africa	Jordan
Norway			Taiwan, China	Kenya
Poland			Thailand	Kuwait
Portugal			Turkey	Lao PDR
Romania			Vietnam	Lebanon
Slovak Republic			Yemen, Rep.	Mongolia
Slovenia			Zimbabwe	Mozambique
Spain				Myanmar
Sweden				Nigeria
United Kingdom				Oman
				Pakistan
				Palau
				Palestine
				Panama
				Peru
				Puerto Rico
				Samoa
				Saudi Arabia
				Sri Lanka
				Tanzania
				Uganda
				United Arab Emirates
				Uzbekistan
				Venezuela, RB

Notes: This table presents countries categorized according to their level of compliance with EU Privacy legislation.

[a] In July 2020, the EU Court of Justice invalidated the the EU-U.S. Privacy Shield Framework. We consider that US belongs to the category *Recognized by the EU*. From July 2020, US does not belong anymore to this category.

Supplementary Appendix C:

Size of Apps and Downloads

To measure the market size of a given app, we use the download category provided by Google Play Store that includes 21 distinct groups. The number of downloads are presented in Table 13 and range from 0 to over five billion downloads. It shows the mean of apps across download intervals.

Table 13: **Summary Statistics: Distribution of Downloads**

	Mean	Min	Max
Download=0	0.0007	0.0	1.0
Download=1	0.0142	0.0	1.0
Download=5	0.0132	0.0	1.0
Download=10	0.0588	0.0	1.0
Download=50	0.0354	0.0	1.0
Download=100	0.0988	0.0	1.0
Download=500	0.0468	0.0	1.0
Download=1k	0.1137	0.0	1.0
Download=5k	0.0504	0.0	1.0
Download=10k	0.1122	0.0	1.0
Download=50k	0.0524	0.0	1.0
Download=100k	0.1363	0.0	1.0
Download=500k	0.0634	0.0	1.0
Download=1000k	0.1246	0.0	1.0
Download=5000k	0.0330	0.0	1.0
Download=10000k	0.0343	0.0	1.0
Download=50000k	0.0053	0.0	1.0
Download=100000k	0.0050	0.0	1.0
Download=500000k	0.0008	0.0	1.0
Download=1000000k	0.0008	0.0	1.0
Download=10000000k	0.0001	0.0	1.0
Observations	1509000		

Notes: The table illustrates the distribution of apps per download range and it indicates the lower range.