Privacy and Children: What drives digital data protection for very young children? PRELIMINARY VERSION

G. Cecere[*], F. Le Guel[†], V.Lefrere[‡], C.Tucker[§], P.L. Yin[¶]

July 10, 2018

Abstract

Internet content and educational apps provide wonderful opportunities for children - a vulnerable group of users which spends huge amounts of time playing on apps. However, children do not always have a good understanding of safety and privacy. Using an original dataset of apps targeted at very young children, we explore the types and scope of data that are collected via children's use of online mobile applications. We show that in the global app developers economy, collection of sensitive data varies with the developer's geographical location. Developers based in the OECD countries (which includes the USA) are less likely to collect sensitive data than developers in countries with no privacy laws. Developers that opt in to an official Google program which encourages compliance with USA child privacy regulation collect fewer sensitive data unless those developers come from countries with weak privacy regulation.

JEL CODE: D82, D83, M31, M37

[*]Institut Mines Telecom, TEM and Digital Society Institute. Email: grazia.cecere@telecom-em.eu

[†]RITM-University of Paris Sud. Email: fabrice.le-guel@u-psud.fr

[‡]Telecom Ecole de Management, Institut Mines Telecom-RITM-University of Paris Sud. Email: vincent.lefrere@u-psud.fr

[§]Massachusetts Institute of Technology (MIT) - Management Science (MS). Email: cetucker@mit.edu

[¶]Greif Center for Entrepreneurial Studies, Marshall School of Business, University of Southern California. Email: pailingy@marshall.usc.edu

We are grateful to Qinhou Li for helpful comments and suggestions. Special thanks are due to the ANR Project for their support. An earlier version of this paper was presented at IIOC Indiana 2018, AFSE Paris 2018 ZEW ICT Conference 2018. Work supported by the DAPCODS/IOTics ANR 2016 project (ANR-16-CE25-0015). We also want to thank our reseacrh assistant for their amazing work namely: Theo Marquis, Hugo Allouard and Enxhi Leka. All errors are the sole responsibility of the authors.

1 Introduction

Many mobile applications are targeted at very young children including toddlers (up to 3-yar olds) and preschool-aged children. According to a recent Common Sense report, 98% of under 8-year olds use mobile devices for an average of 48 minutes per day (Rideout, 2017). While mobile applications offer children great learning opportunities, they can collect user data since children do not have a good grasp of safety and privacy. Similar to adult apps, these child apps automate the collection of detailed data on their users. However, to our knowledge, there are no empirical studies on the extent and influence of the digital data that are collected on child users. Digital platform self-regulation aimed at safeguarding children can help developers to produce apps that can comply with the USA law, and offer parents the opportunity to identify safe and appropriate content for their children. Given the widespread use of mobile applications by children[1], what is it that constrains developers from collecting sensitive data on children?

Both the USA and the EU have specific regulation designed to protect children's privacy: COPPA (Children's Online Privacy Protection Act) and GDPR (General Data Protection Regulation), respectively.[2] In particular, in January 2013[3], the USA Federal Trade Commission (FTC) clarified that in order to collect geo-location data from a child, the app developer must have documented consent from its parent[4]. In theory, the ruling should apply to any app that could be used by a child residing in the USA. Indeed, the FTC has prosecuted the developers of mobile applications targeted at children that collect geo-location data without

[1]The recent Mobile Kids Report published by Nielsen (2017) shows that 59% of the children interviewed use mobile devices to download apps `http://www.nielsen.com/us/en/insights/news/2017/mobile-kids--the-parent-the-child-and-the-smartphone.html`, retrieved January 8, 2018.

[2]Children's Online Privacy Protection Act of 1998, 16 CFR Part 312, www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule (retrieved January 8, 2018) and for the EU, see the newly adopted General Data Protection Regulation; 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, (retrieved March 22, 2018)

[3]See the Children's Online Privacy Protection Rule: https://www.ecfr.gov/cgi-bin/text-idx?SID=cbe35c6ccc2aaf22d50f0087848c30c8&mc=true&node=pt16.1.312&rgn=div5, retrieved January 8, 2018. See also Macenaite and Kosta (2017) for a comparison of EU and USA laws.

[4]More precisely, verifiable parental consent is required for the collection, use, and disclosure of personal information on children aged under 13. Personal information includes geolocation details that are sufficiently precise to identify the street name and city, photos, videos, and audio from a child, screen name or user name, persistent identifiers to recognize an app user over time and across different applications

explicit parental permission.[5]

This paper provides some broad initial data on both compliance with regulation designed to protect children's privacy, and the factors that appear to influence that compliance. We explore whether compliance with laws designed to protect child privacy vary with the age of the child targeted, the developer's location, and the platform initiatives designed to encourage compliance. Specifically, we examine a May 2015 Android initiative where Google Play Store introduced a form of self-regulation called the "Designed for Families" (DfF) program, to encourage developers to comply with COPPA, and to help parents identify content appropriate for children.[6]

We collected weekly data from July to September 2017 on Google Play apps in the DfF category. We compare these apps to apps which did not choose to certify but which targeted children through the use of keywords such as "preschool" and "toddler" in their app descriptions. Our dataset includes 9,799 apps corresponding to 4,442 different developers located in 86 countries, generating a panel of 92,746 observations. To measure the effects developers' country regulation, we identify developer country based on the addresses provided by the developers. Apps in Google Play are automatically released worldwide with automated translation of the app description unless the developer specifies otherwise.

The results show that developers located in regions with no privacy regulations collect more sensitive child data relative to developers based in the USA or the other OECD countries. However, unless they are located in countries with no strong privacy regulation developers who comply with the Google self-certification program are less likely to collect child data. The worldwide distribution of apps means that compliance with DfF induces a children's privacy protection spillover effect across countries. The results are robust to a broader definition of sensitive data and the granularity of the location data. However, it should be underlined that Chinese developers are likely to collect sensitive user data aimed at perfectly identifying the user (IMEI number) but are less likely to collect location data.

[5] See url https://www.ftc.gov/news-events/press-releases/2014/12/ftc-warns-childrens-app-maker-babybus-about-potential-coppa

[6] During Google's 2015 Annual Conference, app developers were introduced to the "Family star" icon. Note also that in 2013, the Apple App Store introduced a kids app category (Apple's WWDC 2013 Keynote).

We contribute to three literature streams: the economics of privacy, the economics of smartphone applications, and the more general literature on children's Internet usage. The literature on the economic effects of privacy regulation highlights a trade-off between protecting the individual and the development of further innovations. Goldfarb and Tucker (2012) focuses on the effects of privacy regulation on firm performance while Campbell *et al.* (2015) examine competition and Miller and Tucker (2009) examines welfare outcomes. To our knowledge, the present study is the first to document the effects of privacy regulation with a focus on protecting children's privacy. It builds on the finding in Rochelandet and Tai (2016) that there is a relationship between privacy regulation and location. We show that in the global app economy, developers are influenced by the existence or lack of regulation, and that there can be international spillovers from privacy regulation on behavior.

Our findings have direct relevance for the second literature stream on the economics of mobile applications. This work focuses mainly on the characteristics of killer apps, and estimation of the demand and supply conditions. Ghose and Han (2014) use a structural model to estimate the factors influencing consumers' demand for apps. Their results suggest that demand for children's apps is higher than demand for adult apps. They show also that kids' apps have lower marginal production costs compared to other age categories. Yin *et al.* (2014) investigate the differences between game and non-game apps in relation to achieving "killer app" status. They find that developers of non-game apps have a higher chance of developing a killer app if they focus on a single app, and improve it via updates. In the case of games apps, the probability of a particular app being successful increases with the developer's experience. We build on the body of work which demonstrates the role played by platform design on the strategies of app developers. Ershov (2017) investigates how the design of the Google Play platform changed the entry dynamics, and shows that splitting games categories into different subcategories reduces search costs and lowers the quality of new entrants. Kummer and Schulte (2016) find a trade-off between app demand and supply interests: the amount of personal information collected to monetize a given app reduces app success measured by the number of downloads. While there is empirical evidence showing the importance of game categories in the smartphone market, there is no published economics and management research on the characteristics of apps aimed at children, and especially work on how platform policy can support regulation and influence individual behavior. The exception is the recent paper by Reyes *et al.* (2018) that analyzes popular free mobile apps. The authors show that the majority of apps do not comply with USA child privacy regulation. Our paper extends this analysis by using a larger, pooled sample, and taking account of both developers' location and the impact of the DfF program on COPPA compliance.

Finally, our research adds to two broader streams of research on children's use of the Internet. One finds that Internet access has mixed effects on education outcomes (Bulman and Fairlie, 2016; Belo *et al.*, 2013); the other studies the relationship between the presence of children in the household and Internet use. There is empirical evidence that Internet use in school affects the level of Internet penetration in households (Belo *et al.*, 2016). We contribute to this literature by highlighting children's participation in the mobile app economy.

This paper has several implications for policy. First, the statistics we provide on the scope and depth of the data collected on children are an improvement on several existing policy studies. Two FTC policy reports (FTC, 2012a,b) provide some initial summary statistics on data collection by apps but evaluate only 364 apps. These studies focus on the extent to which these apps disclose their data collection activity via privacy policies. Another study of websites conducted by the Global Privacy Enforcement Network analyzes the privacy practices of 1,494 world websites targeting children.[7] It finds that 67% of these websites require personal information: 29% asked for names, 20% asked for dates of birth, 12% asked for phone numbers, 11% asked for addresses, and 9% gathered photos or videos (GPEN, 2015). We show that in the mobile applications economy (which is increasingly replacing desktop-orientated websites), collection of data especially on very young children may be even more pervasive. This is because, unlike websites, mobile applications do not rely on the child being able to type or report information but automate its collection, meaning they collect data on very young children in particular.

As well as providing some of the first and most comprehensive information on automated data collection practices related to very young children, our empirical analysis provides evidence that should inform future policy. Second, we identify spillover effects from platform

[7] GPEN includes 29 Data Protection Authorities worldwide - '2015 GPEN Sweep - Children's Privacy': http://194.242.234.211/documents/10160/0/GPEN+Privacy+Sweep+2015.pdf, retrieved January 8, 2018.

compliance efforts surrounding USA policy regulation on the behavior of foreign developers. Third, our analysis suggests that in the global app economy, although some developers are subject to regulation, collection of child data is pervasive in non-regulated countries. Many international developers do not appear to comply with child privacy regulation.

This paper is structured as follows. Section 2 describes the data sources and presents the descriptive statistics. Section 3 presents the econometric models. Section 4 discusses the econometrics results and provides some robustness checks. Section 5 concludes.

2 Description of the sample

We collected weekly data on smartphone applications for children from the USA Google Play Store. Developers who produce child apps can decide to opt-in to DfF, or they can post their apps to the Play Store without complying with the DfF program. First, we collected the characteristics of apps in the DfF program aimed at children aged under 13 years. The DfF program includes six broad categories: Action & Adventure, Brain Games, Creativity, Education, Music and Video, and Pretend Play. The apps indicate appropriate age categories: children aged 5 & under, children aged 6-8 years, children aged 9 years & and over, or mixed audience. For the purposes of our analysis, we collected data on apps targeted at children aged under 13 years which includes the category 9 years & over.[8] Developers who opt in to this program self-declare compliance with COPPA, along with other requirements specified by Google. Apps submitted to DfF are subject to a special review process by Google. Second, we constructed a benchmark group of applications aimed at children by simulating the user's (parent's) likely keyword search process in Google Play to identify child apps. Using the Google Adwords keyword planner tool, we identified the list of keywords most frequently associated to child applications: children, children's, kids, baby, babies, toddlers, educational, toddler, preschool, preschoolers, child monitoring, kindergarten, kindergarteners, boys, girls, kid monitoring, 2 year old, 3 year old, 4 year old, 5 year old, 6 year old, 7 year old, 8 year old, 9 year old, 10 year old, 11 year old, 12 year old.

[8]Appendix figure 5 shows the DfF program menu.

Our sample consists of apps included in the Google Play DfF program plus apps identified at least once by the keyword searches conducted during the period of study. We tracked each application over a period of 12 weeks, starting from its first appearance to the end of the sample period. We excluded apps that appeared only once in our dataset [9]. New apps appear over time while others become unavailable: the number of apps available in the DfF program or identified by the keyword searches increased from 5,137 to 9,799. Our sample includes 92,746 observations; 78.8% of the applications included a clear developer address. A dummy variable, *Presence in the Masterlist*, indicates whether the app is present each week in the DfF program or in the benchmark search. Developers were located in 86 countries. Table 1 provide descriptive statistics of the overall sample. Table 2 presents the breakdown statistics for the categories of downloads .

Our empirical strategy allows us to measure whether the platform policy related to children's content provides effective protection for their personal data compared to the benchmark group. We collected all publicly available data such as app characteristics (e.g. user ratings, freemium, free, paid), developer's name and address, type of interactive elements utilized by the app, and number and type of permissions[10] required by developers.

We are interested in 1) measuring the effectiveness of the platform policy to protect children, and 2) testing whether national privacy regimes correspond to developers' collection of fewer sensitive data.

2.1 Dependent variables: COPPA and LOCATION

To measure whether children's apps comply with USA children privacy legislation we identify permissions that possibly violate to the COPPA regulation. We created the binary variable *COPPA* which considers whether the app collects any sensitive data covered by COPPA regulation. COPPA identifies various types of data such as photos, videos, and audio files that contain children's images or voices as violations of children's privacy. It also covers geolocation data such as street name and city name.[11] *LOCATION* is constructed similar to

[9]We exclude 481 applications that appear only once.
[10]We collect permissions over time.
[11]The complete list of children's personal data is available at https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule.

Table 1: **Summary statistics for the full sample of apps**

	Mean	SD	Min.	Max.	N
COPPA	0.358	-	0	1	92746
LOCATION	0.128	-	0	1	92746
DfF: Less than 5	0.128	-	0	1	92746
DfF: Between 6-8	0.131	-	0	1	92746
DfF: 9 and more	0.152	-	0	1	92746
Log number of reviews	5.318	(3.480)	0	18	92746
User rating	3.820	(1.228)	0	5	92746
Update Before DfF	1.613	(5.181)	0	60	92746
Update After DfF	17.568	(9.979)	0	29	92746
Presence in the masterlist	0.650	-	0	1	92746
Top ranking ratio	0.361	-	0	1	92746
Freemium	0.318	-	0	1	92746
Price	0.799	(2.317)	0	99.990	92746
Users interact	0.048	-	0	1	92746
Contains ad	0.578	-	0	1	92746
DfF	0.722	-	0	1	92746
Without developer address	0.212	-	0	1	92746
OECD	0.340	-	0	1	92746
No OECD	0.171	-	0	1	92746
USA	0.224	-	0	1	92746
China	0.053	-	0	1	92746
Member of the UE	0.264	-	0	1	92746
Recognized by the EU	0.060	-	0	1	92746
Independent authority	0.038	-	0	1	92746
With legislation	0.123	-	0	1	92746
No privacy law	0.025	-	0	1	92746
High income	0.378	-	0	1	92746
Upper middle income	0.055	-	0	1	92746
Low and middle income	0.078	-	0	1	92746
Developer: Downloads	13.033	(5.213)	0	23	92746
Developer: User ratings	3.560	(1.378)	0	5	92746
Developer: No User ratings	0.060	-	0	1	92746
Developer: Missing	0.088	-	0	1	92746
Developer: Before DfF	20.205	(19.07)	0	64	92746
Developer: After DfF	3.325	(7.327)	0	29	92746

Notes: Descriptive statistics of the full sample.

COPPA using a more limited set of data types that only collect location information, thus it takes the value 1 if the app collects location data and 0 otherwise.

Table 13 presents means of the permissions and interactive elements required to construct the dependent variables *COPPA* and *LOCATION*. Column 1 presents the statistics for the whole sample, and columns 2 and 3 present the respective app statistics for USA and EU de-

Table 2: **Summary statistics for the categories of downloads**

	Mean	SD	Min.	Max.	N
Download 0	0.002	(0.043)	0	1	92746
Download 1	0.016	(0.127)	0	1	92746
Download 5	0.011	(0.103)	0	1	92746
Download 10	0.056	(0.229)	0	1	92746
Download 50	0.032	(0.176)	0	1	92746
Download 100	0.091	(0.288)	0	1	92746
Download 500	0.046	(0.210)	0	1	92746
Download 1000	0.115	(0.319)	0	1	92746
Download 5000	0.052	(0.221)	0	1	92746
Download 10000	0.117	(0.321)	0	1	92746
Download 50000	0.059	(0.235)	0	1	92746
Download 100 000	0.138	(0.345)	0	1	92746
Download 500 000	0.067	(0.250)	0	1	92746
Download 1 000 000	0.137	(0.344)	0	1	92746
Download 5 000 000	0.029	(0.167)	0	1	92746
Download 10 000 000	0.027	(0.162)	0	1	92746
Download 50 000 000	0.004	(0.061)	0	1	92746
Download 100 000 000	0.003	(0.054)	0	1	92746
Download 500 000 000	0.000	(0.012)	0	1	92746
Download 1000 000 000	0.000	(0.007)	0	1	92746

Notes: Descriptive statistics of the category of download

velopers, and column 4 presents the statistics for China (including Hong Kong). *LOCATION* is based on *ALEC* (Access Location Extra Commands) used to determine user locations based on various device capabilities, and *ANBL* (Approximate Network Based Location) used to access approximate location derived from network location sources such as cell towers and Wi-Fi. These last two are usually used to display location based ads by publishers such as Admob[12], *MLST* (Mock Location Sources for Testing, used to facilitate developer testing of geolocation data applications), *Precise GPS Location*, and the interactive element *Share Location*. A child's image and voice can be captured via the permissions *Take Pictures and Videos* and *Record Audio*. The permission *Read Phone Status and Identity* allows developers to identify a smartphone's unique IM (International Mobile) identifier.

[12]https://android.izzysoft.de/applists/perms?lang=en (retrieved April 28, 2018).

2.2 App characteristics

Google Play provides a large set of information for all apps (see Table 1). To measure app success, we include the variable *Log Number Reviews*, and also include a set of 20 variables measuring download intensity ranging from 0-10 to more than 1 billion (see Table 2). To measure app popularity, we use the ranking indicated by the user (variable *User Rating*) on app quality, on a 0 to 5 scale.

While many apps are free (about 41.4% of the whole sample) and paid apps represent 26.8% of applications. The variable *Price* indicates the price of the apps going from zero to free apps to 99.99 the maximum price. Finally, the binary variable *Freemium* indicates whether the application offers IAP (in-app purchases). This applies to 31.8% of the applications in the sample. The binary variable *Contains Ad* takes the value 1 if the app displays advertisements to users. Overall, 57.8% of apps include ads.

We employ one of those elements *Users Interact*, as a variable to measure sensitive data collection. It 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.[13] The ESRB (Entertainment Software Rating Board) is a non-profit, self-regulatory body that assigns ratings to video games and apps to classify content according to its target audience. ESRB designates a set of interactive elements that affect the ratings for child-appropriate apps. Google Play Store indicates which of these elements are implemented by the app.

2.3 Geographical location of developers

To explore regulation spillovers to other countries, we retrieved geographical information disclosed by developers of apps available in the Google Play store. First, using Google Maps APIs we collected location latitudes and longitudes to identify the country. Second, we created an algorithm to search for a country name in the developer address [14]. Third, we checked the match between the location identified using the Google Maps APIs and the country name identified by the algorithm. Fourth, we did a manual check for certain

[13] `http://www.esrb.org/ratings/ratings_guide.aspx/#elements`, retrieved January 8, 2018.

[14] We retrieve the geographical address using four different geographical dataset: Google Map, Bing Map, Open stree Map, Geo Map

addresses. Fifth, we identified missing geographical location information and created the variable *Without developer address*. COPPA legislation requires that parents be informed about the companies that collect child data, and in particular, that companies indicate their contact details such as email or geographical location. Although the platform mandates developer's address (since September 30, 2014) for those developers offering paid apps, in-app purchases, and payment through the app, we found that 21.2% of the apps in our sample included no developer address.

USA developers produced 22.4 % of the apps in our sample. Developers in European countries represent 26.4% (the UK was responsible for 5.6 % of the apps). China with Hong Kong accounts for 5.3% of the whole sample.

2.3.1 National privacy regulation

Privacy regulation rules vary across countries, and we exploit this variation to characterize countries' privacy policies. To assess differences in national regulatory frameworks, we augment these data with a vector of the *Institutional framework* measures associated to the developer's address. Since developers in the USA develop apps for their domestic market, it is reasonable to believe that their behavior may differ from that of other developers. Therefore, we created the dummy variable *USA* which measures whether the developer is located in the USA. We also created a dummy variable *China* which takes the value 1 if the developer is located in China or Hong Kong. Appendix Figure 4 presents some empirical evidence showing that the behavior of developers from China and Hong Kong does not differ).

We use a measure of privacy regulation to indicate the country's level of compliance with EU privacy legislation.[15] This index was computed by the French Privacy Regulation Authority (CNIL).[16] The dummy variable *EU* identifies developer country belonging to the EU or the EEA (European Economic Area) and indicates that the country's privacy laws are compatible with EU legislation. Similarly, the variable *Recognized by EU* indicates that the privacy laws in a country outside EU are compatible with EU privacy laws. The binary variable *Independent authority* indicates the existence of an independent privacy regulation

[15] Appendix table 11 indicates the countries belonging to each privacy legislation group.

[16] `https://www.cnil.fr/fr/la-protection-des-donnees-dans-le-monde`, retrieved January 8, 2018.

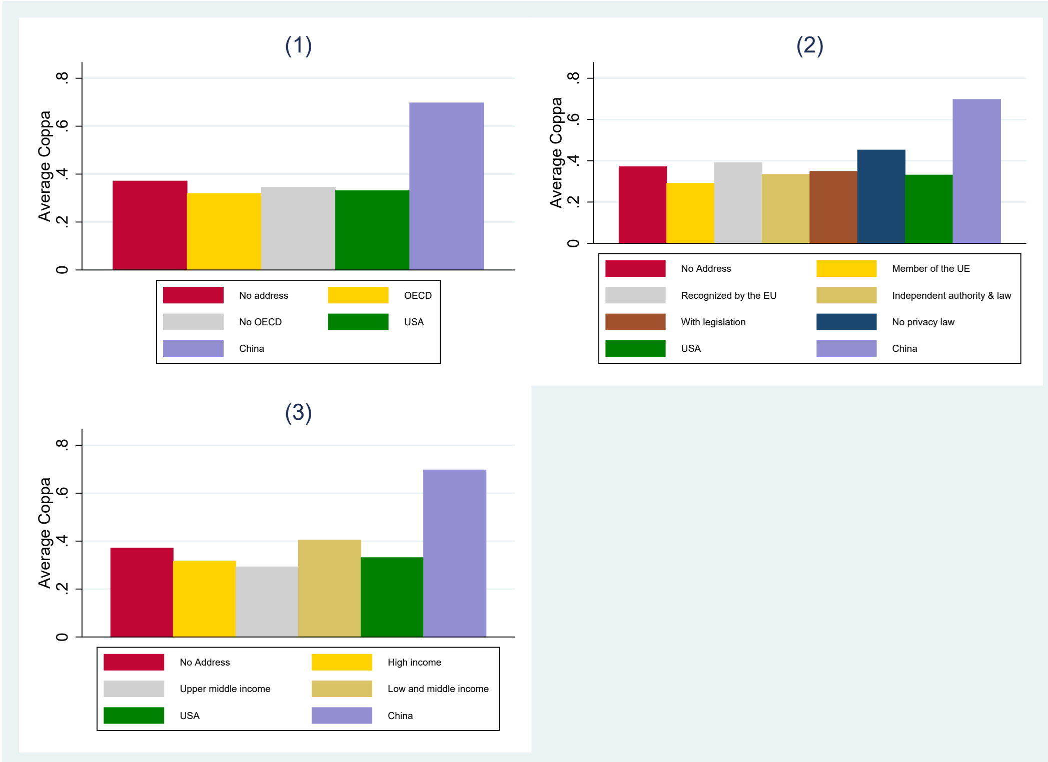
authority in the app developer's country and the presence of a privacy legislation framework. The binary variable *With Legislation* indicates that the country has privacy legislation only (but no independent authority regulating privacy), and the dummy variable *No Privacy Law* indicates the absence of privacy laws in the developer's country (2.5% of apps). Appendix table 11 presents countries categorized according to their level of compliance with EU privacy legislation.

The developer's strategy might also be associated to the home institutional framework. To measure this effect, we include two sets of variables which exclude *USA* and *China*. First, we consider whether OECD country (excluding USA) developers demonstrate behavior that is different from that displayed by developers located in non-OECD countries (excluding China) that have weaker institutions and regulation. Second, we include the country income level computed by the World Bank, to measure the effect of the developer's origin country's economic growth. The vector of the income variables includes *High Income, Upper and Middle Income, and Low and Middle Income.* This set of variables proxies for the relative costs associated to the collection and storage of personal data across countries.

2.3.2 Graphical evidence

Figure 1 depicts the percentage of apps per country group that collect children's personal data. The graphical evidence shows that overall, developers located in China collect more data compared to developers in the USA and other country groups. The histogram (1) in figure 1 shows the distribution of sensitive data items for OECD and non-OECD countries, with USA and China separated out from the group of countries. Developers in China collect more sensitive data compared to all other locations, followed by developers who do not list an address. Histogram (2) in figure 1 shows the distribution of the number of sensitive data items collected according to the EU privacy regulation regime, again with USA and China separated out. Developers in China collect more sensitive data compared to all other locations, followed by developers in countries with no privacy laws. Histogram (3) figure 1 depicts the distribution of sensitive data according to level of income. Developers in China collect more sensitive data, followed by developers from low and middle income countries.

Figure 1: **Collection of COPPA sensitive data by developer location characteristics**



Notes: The vertical axis is the percentage of apps collecting COPPA sensitive data.

2.4 Sensitive data and users' location data by age group

DfF measures whether the app complies with the DfF program related to apps for children aged under 13 years. The program includes three age groups which we distinguish using three sets of data: *Under 5 years, Between 6-8 years, and 9 years and over.* Table 3 presents the percentage of apps that collect at least one piece of sensitive user data, by age group. Column 1 shows that in the whole sample 10.6% of apps that are targeted at children aged 5 & under that certify themselves via the Self-Regulatory Regime program collect at least one piece of sensitive data. Developers in Europe are less likely to collect data, especially those in the benchmark category. Column 4 shows that in the self-regulated regime, developers in China tend to collect more data than other developers.

Table 3: **Percentage of Applications that request at least one COPPA sensitive data per age group**

	(1) Overall	(2) USA	(3) EU	(4) China
DfF: Less than 5	0.106	0.161	0.129	0.176
DfF: Between 6-8	0.110	0.135	0.113	0.152
DfF: 9 and more	0.162	0.175	0.173	0.316

Notes: Column 1 presents the percentage of apps within each age group that collected at least one piece of COPPA identified sensitive data. Column 2 restricts apps in Column 1 to those developed in the USA, Column 3 restricts apps to those developed in the EU, and Column 4 restricts apps to those developed in China (including Hong Kong).

2.5 Developer characteristics

In addition to developer's address, using the AppBrain website we collected data to measure developer characteristics. In particular, it is interesting to differentiate professional and experience developers. These data include the overall rating associated to developers, the overall number of installation, and the date of entry in Google Playstore. This set of variables permits to measure also whether the collection of personal data is associated with huge amount of data collected.

3 Model specification

Our econometric analysis estimates the effect of the regulation in the developer's country of origin on the among of child sensitive data collected. Our dependent variable is a binary variable, and we use a probit estimation with robust standard errors clustered on the app level. We model the probability to collect sensitive data using the following specification:

$$\mathbf{P}(\mathbf{COPPA}) = \alpha_0 + \mathbf{X}\beta + \mathbf{Z}\gamma + \mathbf{D}\theta + \rho_t + \epsilon \tag{1}$$

Our primary vectors of interest are country privacy regulation, $\mathbf{X}$. We include the level of privacy protection according to European legislation using a set of dummy variables which capture the country's level of compliance. We consider also whether the developer is located in an OECD country. We include income level as reported by the World Bank, and include dummies for developers from *China* (incl. *Hong Kong*) and *USA*. $\mathbf{Z}$ is the vector of app characteristics i at time t developed in country j. $\mathbf{D}$ is the vector of developers characteristics i at time t developed in country j. ϵ is an independent and identically distributed random error term. The equation also includes time (week) effects ρ_t. Since we have very few time varying variables we estimate the model using pooled cross section data.

4 Estimation of the probability to collect sensitive data

4.1 Baseline model: Effect of DfF and developer home country regulation

Table 4 presents the marginal effects of the probit estimations. COPPA legislation precisely defines the sensitive data covered by the legislation, and requires that each company or the third parties that collect user data provide information such as name and address to allow parents to contact them. We investigate the impact of privacy regulation and macro-economic characteristics on the number of items of sensitive data requested by developers. All regressions include app characteristics to control for observable differences in the apps underlying the probability of collecting sensitive data. We also include developer characteristics in order to measure developer experience and size. By including the full set of app characteristics in these models, we can abstract away from app heterogeneity. All the specifications include country-level controls and time fixed effects. The omitted category for the institutional variables is developers that do not indicate their address. Standard errors are clustered at app level.

The negative and significant marginal effect of the DfF program coefficient suggests that the likelihood of collecting data within the system decreases with participation in the program. This finding is aligned to the intention of the platform to encourage compliance with COPPA legislation. Developers that decide to comply with the DfF program show a 16.3% to 16.9% lower probability of collecting users' sensitive data. The effect of developer location in China is associated to a higher probability (35.3%, column 4, to 35.6%, column 1) (relative to developers with no address) of collecting users' data.

We investigate how DfF moderates the effect of developer country to measure USA legislation spillovers. Column 2 in Table 4 adds a set of interaction terms between developer country and DfF. Unlike China, USA developers are less likely to collect sensitive data relative to developers without an address. Column 3 in Table 4 includes measures of institutional privacy regulation. Column 4 includes a measure of country income developed by the World Bank. Developers in *Low and Middle income* countries are likely to require more sensitive data compared to developers that do not disclose their geographical address.

Table 4: **COPPA sensitive data**

	Probit (marginal effect)			
	(1)	(2)	(3)	(4)
OECD	0.009	0.004		
	(0.015)	(0.023)		
No OECD	0.032**	0.011		
	(0.016)	(0.025)		
USA	0.031*	0.134***	0.029*	0.028
	(0.017)	(0.028)	(0.017)	(0.017)
China	0.356***	0.021	0.354***	0.353***
	(0.024)	(0.061)	(0.024)	(0.025)
DfF	-0.169***	-0.163***	-0.166***	-0.165***
	(0.013)	(0.022)	(0.013)	(0.013)
USA X DfF		-0.124***		
		(0.029)		
China X DfF		0.377***		
		(0.061)		
OECD X DfF		0.005		
		(0.028)		
No OECD X DfF		0.032		
		(0.031)		
Member of the UE			-0.014	
			(0.016)	
Recognized by the EU			0.059**	
			(0.025)	
Independent authority			0.039	
			(0.028)	
With legislation			0.026	
			(0.018)	
No privacy law			0.106***	
			(0.032)	
High income				0.006
				(0.015)
Upper middle income				-0.013
				(0.023)
Low and middle income				0.072***
				(0.021)
Developer characteristics	Yes	Yes	Yes	Yes
App characteristics	Yes	Yes	Yes	Yes
Week fixed effect	Yes	Yes	Yes	Yes
Observations	92746	92746	92746	92746

Notes: Marginal effects of the Probit estimations. The dependent variable is dummy variable COPPA. Robust standard errors clustered at app level reported in parentheses. Omitted category is *No developer address*. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

4.2 Collection of data on very young children

We check our results using alternative subsamples, and conduct falsification tests. We investigate whether the app's age category has an effect on the probability of compliance with COPPA legislation Table 5. Overall, apps targeted at very young children seem to collect the same amount of sensitive data as the benchmark group of applications. To address the concern that the probability to collect data changes within the DfF category, we include in the regression the three age categories proposed in DfF. The probability to collect data from different age categories can change among countries. To address this, we split our sample into four groups according to developers' origin.

Column 1 in Table 5 estimates the probability of collecting sensitive data including the OECD group variables, and the DfF age groups. Column 2 estimates the model for the subsample of apps produced in the USA. Developers in the USA are likely to collect fewer data overall. Column 3 estimates the model for the subsample of apps produced in China. The regressions show that developers in China are likely to collect data on children aged 9 and over. Column 4 estimates the model for the apps produced in Europe. We find that European developers are more likely to collect data in apps targeting under-5 and children aged 9 and over , but they are less likely to collect data of children between 6-8. Conversely, the results in Table 5 column 5 which includes only apps produced in the rest of the world, show that developers in the rest of the world are less likely to collect data in apps targeting children under 5 and children over 9 years, and they are more likely to collect data on children aged between 6-8 years.

Table 5: **Estimations of COPPA sensitive data with app's age categories**

	Probit (marginal effect)				
	(1)	(2)	(3)	(4)	(5)
	Overall	USA	(China)	(Europe)	(Other Country)
OECD	-0.040***				
	(0.015)				
No OECD	-0.001				
	(0.016)				
USA	-0.023				
	(0.017)				
China	0.299***				
	(0.025)				
DfF: Less than 5	-0.034**	-0.013	-0.014	0.023**	-0.144***
	(0.016)	(0.010)	(0.016)	(0.011)	(0.023)
DfF: Between 6-8	-0.039***	-0.111***	-0.091***	-0.027***	0.092***
	(0.014)	(0.008)	(0.017)	(0.010)	(0.026)
DfF: 9 and more	-0.010	-0.088***	0.052***	0.036***	-0.041**
	(0.013)	(0.009)	(0.017)	(0.010)	(0.020)
Developer characteristics	Yes	Yes	Yes	Yes	Yes
App characteristics	Yes	Yes	Yes	Yes	Yes
Week fixed effect	Yes	Yes	Yes	Yes	Yes
Observations	92746	20814	4931	24378	42623

Notes: Marginal effects of the Probit estimations. The dependent variable is dummy variable COPPA. Robust standard errors clustered at app level reported in parentheses. Omitted category is *No developer address*. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

4.3 Robustness check: Regression excluding apps without developer address

Our falsification check addresses the concern that our results might be influenced by apps without a developer address. Table 6 reports the estimation of the equation (1) excluding apps with no developer address. While this sample is likely to suffer from self-selection since the subset of developers that declare their address are likely to be more respectful of platform rules, the results are reassuring since they do not change in a major way. Similarly, countries without privacy laws , and particularly China request more sensitive data. We corroborate previous results related to the effect of DfF which negatively affects the probability to collect sensitive data. Overall, China and USA developers are more likely to collect sensitive data compared to developers in OECD. Column 1 includes the group of variable OECD and the reference group is the group of developers in OECD. Column 2 adds the set of variable interaction terms between DfF and the group of OECD countries. Column 3 includes the set of variable measuring privacy regulation. The reference group is the group of developers in Europe. The last column adds the set of income variable. The reference group is the country with High income.

Table 6: **COPPA sensitive data Without missing address**

	Probit (marginal effect)			
	(1)	(2)	(3)	(4)
No OECD	0.014	-0.005		
	(0.015)	(0.026)		
USA	0.033**	0.140***	0.054***	0.032**
	(0.014)	(0.028)	(0.015)	(0.014)
China	0.341***	0.016	0.363***	0.341***
	(0.024)	(0.060)	(0.023)	(0.023)
DfF RC	-0.177***	-0.164***	-0.174***	-0.173***
	(0.015)	(0.022)	(0.016)	(0.015)
USA X DfF RC		-0.132***		
		(0.029)		
China X DfF RC		0.366***		
		(0.062)		
No OECD X DfF RC		0.032		
		(0.031)		
Recognized by the EU			0.076***	
			(0.023)	
Independent authority			0.058**	
			(0.027)	
With legislation			0.031*	
			(0.017)	
No privacy law			0.098***	
			(0.031)	
Upper middle income				-0.024
				(0.022)
Low and middle income				0.052***
				(0.020)
Developer characteristics	Yes	Yes	Yes	Yes
App characteristics	Yes	Yes	Yes	Yes
Week fixed effect	Yes	Yes	Yes	Yes
Observations	73031	73031	73031	73031

Notes: Marginal effects of the Probit estimations. The dependent variable is dummy variable COPPA. Robust standard errors clustered at app level reported in parentheses. Subsample of developers with an address. Omitted category is *OECD* or *Europe*. Significance levels: $p < .10$, $*p < .05$, $***p < .01$

4.4 Robustness check: Number of Users' location data

We show the robustness of our results to an alternative dependent variable: *LOCATION*. We use the same empirical approach as in Table 4. If there were unobserved heterogeneity issues related to developer without geographical address, we would expect to see different result for these estimations. However, we corroborate previous results. Across all specifications, the decision to collect location sensitive data appears to decrease by the participation to the DfF. Importantly, developers in China are less likely to collect location data. But the coefficient for this is negative and small, and generally insignificant. This corroborates the intuition of the previous statistical evidences. Overall, developers in China are less likely to collect location data as they are likely to require *Read and Phone status* which permits to collect the unique identifier of a smartphone. Table 7 in Column 1 adds the vector of the variables measuring the OECD institutional framework.

Column 2 adds the interaction terms between OECD institutional variables and DfF. The interaction terms *USA x DfF* and *OECD x DfF* are negative and statistically significant suggesting that within the DfF program developers in USA and OECD countries are less likely to collect personal data. Column 3 includes a set of dummies measuring compliance with EU legislation. Developers in EU countries or countries whose privacy laws are compatible with EU legislation request less user location data. Developers in countries with *No privacy law* are more likely to collect location data. In the estimations for privacy legislation in specific countries, this variable might capture underlying effects such as infrastructure or wealth. We address this in column 4 which estimates the model including a set of variables measuring the country's income level according to the World Bank. The coefficients suggest that high income developer countries are less likely to collect location data.

Table 7: **Location sensitive data**

	Probit (marginal effect)			
	(1)	(2)	(3)	(4)
OECD	-0.052***	-0.039***		
	(0.009)	(0.013)		
No OECD	0.014	-0.031**		
	(0.011)	(0.013)		
USA	-0.015	0.011	-0.015	-0.012
	(0.011)	(0.017)	(0.011)	(0.011)
China	-0.032**	-0.068***	-0.031**	-0.031**
	(0.014)	(0.026)	(0.014)	(0.014)
DfF RC	-0.130***	-0.139***	-0.129***	-0.130***
	(0.010)	(0.017)	(0.010)	(0.010)
USA X DfF RC		-0.033*		
		(0.018)		
China X DfF RC		0.078		
		(0.059)		
OECD X DfF RC		-0.016		
		(0.018)		
No OECD X DfF RC		0.089***		
		(0.025)		
Member of the UE			-0.036***	
			(0.010)	
Recognized by the EU			-0.044***	
			(0.013)	
Independent authority			-0.046***	
			(0.015)	
With legislation			-0.017*	
			(0.010)	
No privacy law			0.096***	
			(0.025)	
High income				-0.034***
				(0.010)
Upper middle income				-0.030**
				(0.013)
Low and middle income				0.013
				(0.013)
Developer characteristics	Yes	Yes	Yes	Yes
App characteristics	Yes	Yes	Yes	Yes
Week fixed effect	Yes	Yes	Yes	Yes
Observations	92746	92746	92746	92746

Notes: Marginal effects of the Probit estimations. The dependent variable is dummy variable *LOCATION*. Robust standard errors clustered at app level reported in parentheses. Omitted category is *No developer address*. Significance levels: $p < .10$, $*p < .05$, $* * *p < .01$

5 Third parties and targeted ad

COPPA legislation also regulates the distribution of targeted ads. To identify whether the apps offered targeted ads we identify the third parties that allow their distribution. For this purpose, we collected the data on third parties and we categorized them. "Privacygrade" is a computer science research project that lists third parties that provide targeted ads via app [17].

To measure whether the probability to include targeted ads in apps is associated with the collection of sensitive user data, we constructed *Coppa targeting* which takes the value 1 if the app collects sensitive data (as defined by COPPA) and has third parties that offer targeted ads. Figure 2 presents the percentage of apps that offer targeted ads per developer origin. Figure 3 shows the percentage of apps that offer targeted ads and require sensitive data.

Tables 9 and 10 corroborate previously findings. Overall, the variable DfF is negative in all specifications, suggesting that platform design aiming to ensure the non violation of the COPPA and personalized ad is effectively. The positive and significant coefficients on developers in China, reported in columns 1-4 suggest that these developers collect data to offer targeted ad exept when LOCATION is considered.

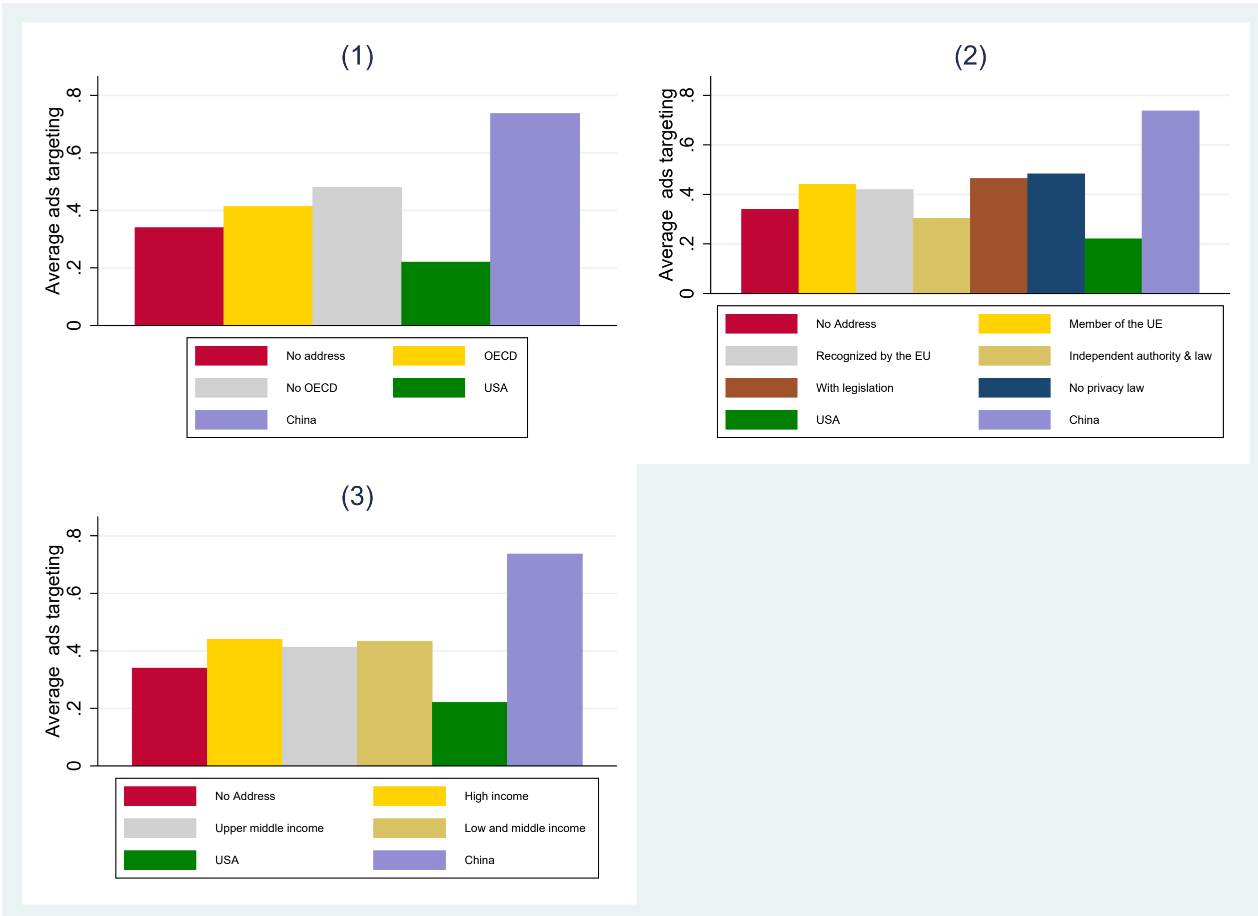
Table 8: **Percentage of Advertising Thirds parties by country**

	(1) COPPA	(2) USA	(3) EU	(4) China
AdMob	0.430	0.189	0.370	0.674
Unity Ads	0.143		0.157	0.258
Chartboost	0.148		0.184	0.151
AppLovin	0.118		0.132	0.244
Vungle	0.108			0.265
Umeng				0.567
AdColony				0.106
Adjust				0.177
HeyZap				0.134
AppNext				0.126

Notes: Distribution of Third parties by country

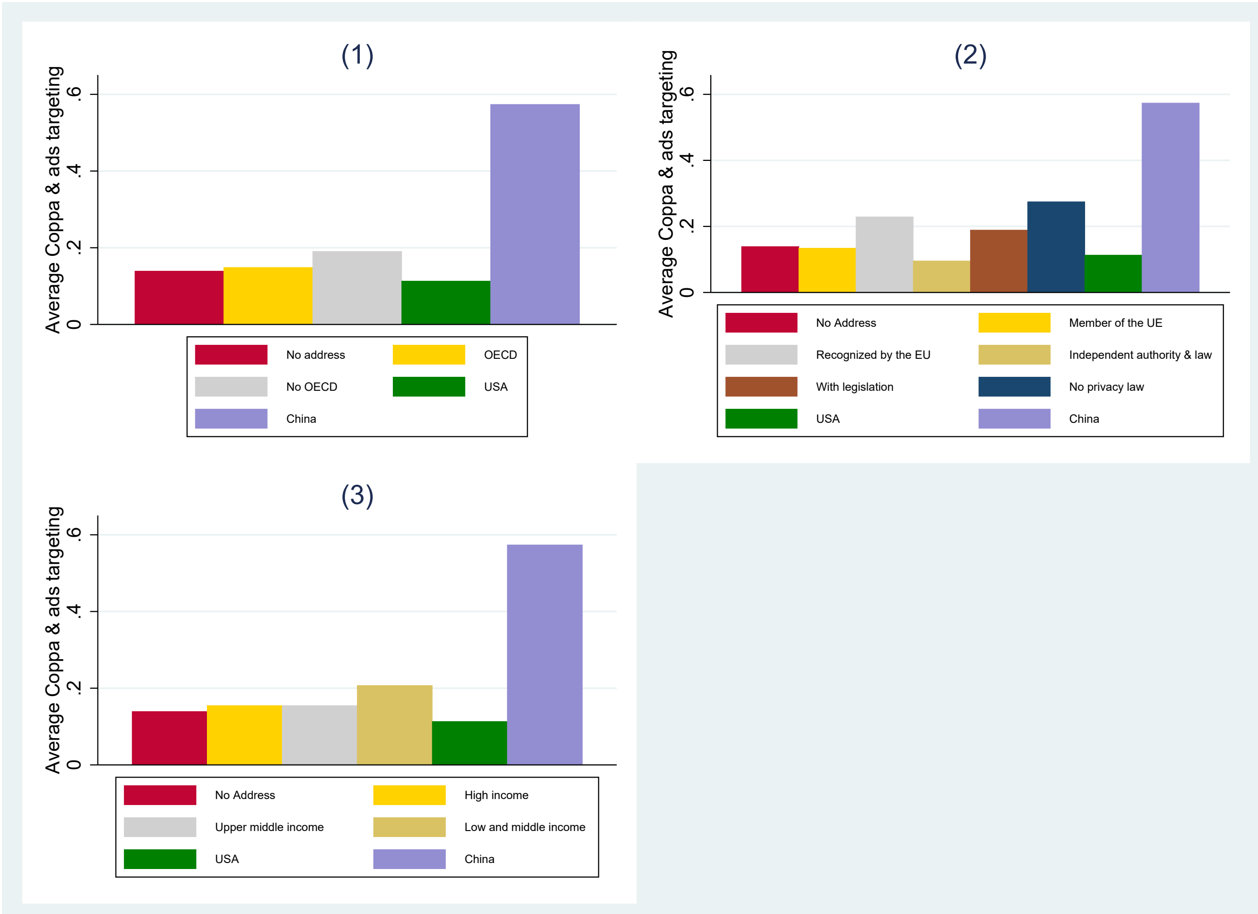
[17] The third parties providing targeted ad are: AdMarvel, AdMob, AdsMogo, AdWhirl, AirPush, AppLovin, CaulyAds, Chartboost, InMobi, Inneractive, Jumptap, LeadBolt, Madhouse SmartMAD, MdotM, Mediba Admaker, Millennial Media, MobClix, MobFox, MobWIN, MoPub, Nexage, Noqoush AdFalcon, Revmob, Smaato, Smart AdServer, Sponsorpay, Tap for Tap, Tapjoy, YuMe

Figure 2: **Percentage of targeted ads per group of countries**



Notes: The vertical axis is the percentage of apps that include targeted ads.

Figure 3: **Distribution of ads targeting & COPPA per group of countries**



Notes: The vertical axis is the percentage of apps that collect sensitive data and use third parties enabling targeted ad.

Table 9: **COPPA and adtargeting sensitive data**

	Probit (marginal effect)			
	(1)	(2)	(3)	(4)
OECD	0.014	0.015		
	(0.012)	(0.015)		
No OECD	0.060***	0.070***		
	(0.013)	(0.018)		
USA	0.021	0.061***	0.020	0.019
	(0.014)	(0.019)	(0.014)	(0.014)
China	0.279***	0.085*	0.280***	0.278***
	(0.022)	(0.044)	(0.022)	(0.022)
DfF	-0.086***	-0.072***	-0.083***	-0.084***
	(0.009)	(0.017)	(0.009)	(0.009)
USA X DfF		-0.059***		
		(0.020)		
China X DfF		0.188***		

		(0.055)		
OECD X DfF		-0.006		
		(0.020)		
No OECD X DfF		-0.018		
		(0.021)		
Member of the UE			0.006	
			(0.012)	
Recognized by the EU			0.067***	
			(0.020)	
Independent authority			-0.022	
			(0.020)	
With legislation			0.055***	
			(0.014)	
No privacy law			0.105***	
			(0.024)	
High income				0.017
				(0.011)
Upper middle income				0.039**
				(0.019)
Low and middle income				0.077***
				(0.017)
Developer characteristics	Yes	Yes	Yes	Yes
App characteristics	Yes	Yes	Yes	Yes
Week fixed effect	Yes	Yes	Yes	Yes
Observations	92746	92746	92746	92746

Notes: Marginal effects of the Probit estimations. The dependent variable is dummy variable Location and targetting advetising thirds parties. Robust standard errors clustered at app level reported in parentheses. Omitted category is *No developer address*. Significance levels: $p < .10$, $*p < .05$, $* * *p < .01$

Table 10: **Location and adtargeting**

	Probit (marginal effect)			
	(1)	(2)	(3)	(4)
OECD	-0.013*	-0.007		
	(0.007)	(0.009)		
No OECD	0.036***	0.017		
	(0.009)	(0.011)		
USA	0.011	0.018	0.011	0.014
	(0.009)	(0.012)	(0.009)	(0.010)
China	0.010	0.002	0.012	0.013
	(0.012)	(0.024)	(0.012)	(0.012)
DfF	-0.048***	-0.052***	-0.047***	-0.048***
	(0.006)	(0.013)	(0.006)	(0.006)
USA X DfF		-0.009		
		(0.014)		
China X DfF		0.013		
		(0.030)		
OECD X DfF		-0.008		
		(0.014)		
No OECD X DfF		0.035*		
		(0.019)		
Member of the UE			0.003	
			(0.008)	
Recognized by the EU			-0.017	
			(0.010)	
Independent authority			-0.042***	
			(0.008)	
With legislation			0.014	
			(0.009)	
No privacy law			0.093***	
			(0.022)	
High income				0.002
				(0.007)
Upper middle income				0.005
				(0.012)
Low and middle income				0.032***
				(0.012)
Developer characteristics	Yes	Yes	Yes	Yes
App characteristics	Yes	Yes	Yes	Yes
Week fixed effect	Yes	Yes	Yes	Yes
Observations	92746	92746	92746	92746

Notes: Marginal effects of the Probit estimations. The dependent variable is dummy variable Location and targetting advetising thirds parties. Robust standard errors clustered at app level reported in parentheses. Omitted category is *No developer address*. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

6 Discussion

(FTC, 2018)

7 Conclusion

We investigate whether developer's location affects the amount of sensitive data collected. We rely on original data from Google Playstore, collected using keywords associated with child applications. The content included in the category Designed for Families must comply with Google's guidelines for age-appropriate content and advertising and comply more closely to COPPA.

We find that developers from countries with weak privacy regulation collect more sensitive data. For example, our results show that developers from OECD countries (including the USA) and EU countries tend to comply with COPPA compared to non-member countries. We observe that national income has no impact on the app's intrusiveness. Together, these findings confirm that home country privacy regulation has an impact on the developers privacy behaviors. USA regulation is likely to have an impact on foreign developers if they comply with the DfF program.

We observe that disclosing country location has an impact on the amount of user data collected. More precisely, developers who do not reveal their geographic location are the most likely to collect sensitive data on children. This is an important result from a policy perspective. For instance, the platform might make provision of an address a condition for approval, which could affect the collection of children's personal data.

It is reassuring that Google's privacy policy - via DfF – is effective for encouraging developers to request fewer pieces of sensitive data. The self-regulation of platforms could reinforce the Children's Online Privacy Protection Act, especially for apps targeting very young children. Spillover effects from platform compliance surrounding USA policy regulation are reduced if we consider developers in non-OECD countries. Overall, our results suggest that the child apps market does not respect children's personal data, and that data can be transferred to other countries outside the USA market where there is an absence of privacy regulation resulting in a loss of control over the use of children's data.

References

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.

Bulman, G. and Fairlie, R. W. (2016). *Technology and education: Computers, software, and the internet*. Working Paper 22237. National Bureau of Economic Research.

Campbell, J., Goldfarb, A. and Tucker, C. (2015). Privacy regulation and market structure. *Journal of Economics & Management Strategy*. 24(1), 47–73.

Ershov, D. (2017). *The Effect of Consumer Search Costs on Entry and Quality in the Mobile App Market*. Working Paper.

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.

FTC (2018). *FTC Warns Gator Group, Tinitell that Online Services Might Violate COPPA*. 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 Tucker, C. (2012). Privacy and innovation. *Innovation policy and the economy*. 12(1), 65–90.

GPEN (2015). *2015 GPEN Sweep - Children's Privacy*. Technical report.

Kummer, M. and Schulte, P. (2016). *When private information settles the bill: Money and privacy in Google's market for smartphone applications*. Working Paper.

Macenaite, M. and Kosta, E. (2017). Consent for processing children's personal data in the EU: following in US footsteps? *Information & Communications Technology Law*. 26(2), 146–197.

Miller, A. R. and Tucker, A. R. (2009). Privacy Protection and Technology Diffusion: The Case of Electronic Medical Records. *Management Science*. 55(7), 1077–1093. doi:10.1287/mnsc.1090.1014.

Nielsen (2017). *Mobile kids: the parent, the child and the smartphone*. Technical report. Last seen: January 2017.

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*. 2018(3), 63–83.

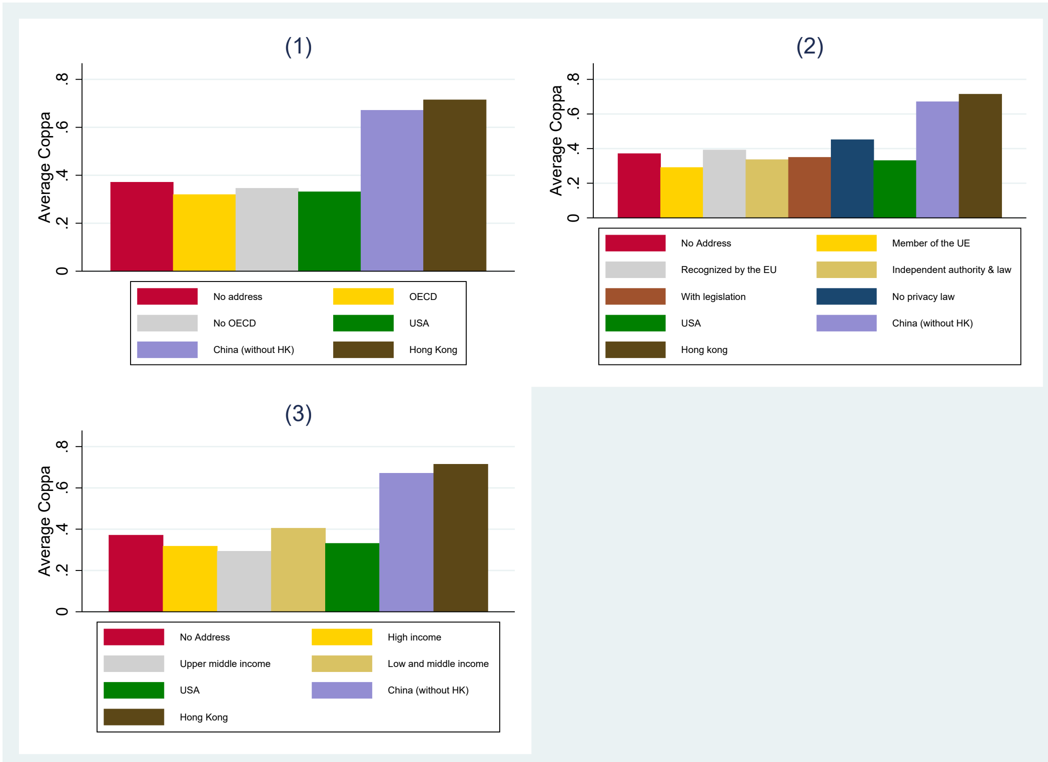
Rideout, V. (2017). The Common Sense census: Media use by kids age zero to eight. *San Francisco, CA: Common Sense Media*, 263–283.

Rochelandet, F. and Tai, S. H. T. (2016). Do privacy laws affect the location decisions of internet firms? Evidence for privacy havens. *European Journal of Law and Economics*. 42(2).

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.

8 Appendix

Figure 4: **Collection of COPPA sensitive data by developer location characteristics with Hong Kong**



Notes: The vertical axis is the percentage of apps collecting COPPA sensitive data.

Figure 5: **Screenshot of Google Play Family**

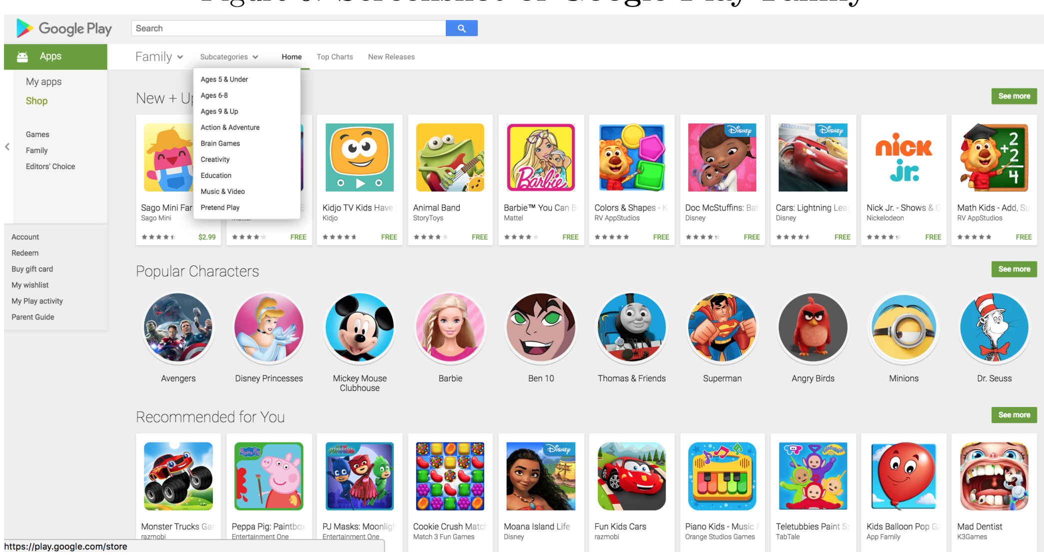


Table 11: **Country names for each Compliance with EU privacy regulation group**

EU	**Recognized by EU**	**Independent Authority**	**With Legislation**	**No Privacy Law**
Austria	Argentina	Armenia	Australia	Bahrain
Belgium	Canada	Azerbaijan	Colombia	Bangladesh
Bulgaria	Israel	Brazil	Hong Kong SAR, China	Belarus
Croatia	New Zealand	Chile	Korea, Rep.	Cambodia
Cyprus	Switzerland	China	Macedonia, FYR	Ecuador
Czech Rep.	United States	Dominican Rep.	Mexico	Egypt, Arab Rep.
Denmark	Uruguay	India	Moldova	El Salvador
Estonia		Indonesia	Morocco	Jordan
Finland		Japan	Nicaragua	Kuwait
France		Kazakhstan	Serbia	Nigeria
Germany		Kosovo	Tunisia	Oman
Greece		Malaysia	Ukraine	Pakistan
Hungary		Philippines	Taiwan	Peru
Iceland		Qatar		Puerto Rico
Ireland		Russian Federation		Saudi Arabia
Italy		Singapore		Sri Lanka
Latvia		South Africa		United Arab Emirates
Lithuania		Thailand		
Malta		Turkey		
Netherlands		Vietnam		
Norway				
Poland				
Portugal				
Romania				
Slovak Rep.				
Slovenia				
Spain				
Sweden				
United Kingdom				

Table 12: DfF subcategories and Keywords

Designed for Families	Keywords		
Ages 5 & Under	2 year old	12 year old	babies
Ages 6-8	3 year old	child	kindergarten
Ages 9 & Up	5 year old	children	kindergartners
Action & Adventure	6 year old	kids	preschool
Brain Games	7 year old	boys	preschoolers
Creativity	8 year old	girls	monitoring
Education	9 year old	toddler	
Music & Video	10 year old	toddlers	
Pretend Play	11 year old	baby	

Table 13: **List of permissions and interactive elements used to construct the two dependent variables *COPPA* and *LOCATION***

LOC.	COPPA	Permissions & Interactive element	(1) Overall Mean	(2) USA Mean	(3) EU Mean	(4) China Mean
✓	✓	ALEC	0.004	0.001	0.002	-
✓	✓	ANBL	0.104	0.085	0.070	0.059
✓	✓	MLST	0.001	0.001	0.000	-
✓	✓	Precise GPS Location	0.088	0.066	0.072	0.041
✓	✓	Shares Location	0.015	0.019	0.011	-
	✓	Read Phone Status And Identity	0.245	0.207	0.198	0.614
	✓	Take Pictures And Videos	0.075	0.090	0.060	0.093
	✓	Record Audio	0.069	0.092	0.056	0.037

Notes: This table depicts the summary statistics of the permissions and interactive elements used to construct the two main dependent variables: COPPA and LOCATION. Column 1 presents the proportion of apps collecting data for the overall sample. Column 2 presents the proportion of apps developed in the USA collecting data. Column 3 presents the proportion of apps developed in the EU collecting data. Column 4 presents the proportion of apps developed in China (including Hong Kong) collecting data.

Table 14: **Percentage of Advertising Thirds parties by country**

	(1) COPPA	(2) USA	(3) EU	(4) China
Precise Gps Location	0.246	0.066	0.072	0.041
Read Phone Status And Identity	0.685	0.207	0.198	0.614
Read Your Contacts	0.060			
Record Audio	0.194	0.092	0.056	0.037

Notes:

Table 15: **Percentage of Social Thirds parties by country**

	(1) COPPA	(2) USA	(3) EU	(4) China
Facebook	0.185	0.119	0.106	0.283
Google Play Games Services	0.062		0.057	0.085
HeyZap				0.134

Notes:

Table 16: **Distribution per Period**

	Mean			
	COPPA	DfF all	Dff age	by category
Week 1	0.588	0.733	0.545	0.526
Week 2	0.570	0.737	0.568	0.465
Week 3	0.626	0.715	0.553	0.311
Week 4	0.626	0.721	0.565	0.294
Week 5	0.622	0.714	0.566	0.276
Week 6	0.609	0.709	0.567	0.351
Week 7	0.589	0.716	0.579	0.324
Week 8	0.582	0.723	0.591	0.287
Week 9	0.587	0.730	0.598	0.298
Week 10	0.586	0.733	0.604	0.254
Week 11	0.601	0.715	0.586	0.231
Week 12	0.627	0.715	0.590	0.264

Table 17: **Percentage of Advertising Thirds parties by country**

	(1) under 5	(2) 6-8	(3) 9 and upper
AdMob	0.171	0.187	0.313
Chartboost	0.161	0.180	0.128
AppLovin	0.106	0.115	0.123
Unity_Ads_mean			0.187
Vungle			0.121

Notes:

Table 18: **Correlations (part 1)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
COPPA (1)	1								
LOCATION (2)	0.51	1							
DfF RC (3)	-0.17	-0.22	1						
Log number of reviews (4)	0.16	0.043	-0.21	1					
User rating (5)	0.016	0.0033	-0.023	0.35	1				
presence (6)	0.013	-0.0042	-0.051	0.31	0.13	1			
change download (7)	-0.012	0.029	0.16	-0.28	-0.070	-0.13	1		
topranking_ratio (8)	0.014	-0.024	0.048	0.39	0.15	0.76	-0.087	1	
Freemium (9)	0.032	0.044	0.040	0.36	0.11	0.12	0.025	0.15	1

Table 19: **Correlations (part 2)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
COPPA (1)	1								
LOCATION (2)	0.51	1							
DfF (3)	-0.17	-0.22	1						
Without developer address (4)	0.015	0.12	-0.30	1					
OECD (5)	-0.057	-0.11	0.12	-0.37	1				
No OECD (6)	-0.011	0.071	-0.014	-0.24	-0.33	1			
USA (7)	-0.029	-0.038	0.12	-0.28	-0.39	-0.24	1		
China (8)	0.17	-0.034	0.087	-0.12	-0.17	-0.11	-0.13	1	
Member of the UE (9)	-0.083	-0.068	0.13	-0.31	0.69	-0.094	-0.32	-0.14	1
Recognized by the EU (10)	0.018	-0.033	0.049	-0.13	0.33	-0.088	-0.14	-0.060	-0.15
Independent authority (11)	-0.0096	-0.025	0.045	-0.10	0.16	0.052	-0.11	-0.047	-0.12
With legislation (12)	-0.0066	0.016	-0.073	-0.19	-0.14	0.66	-0.20	-0.089	-0.22

Table 20: **Summary statistics for the full sample of apps versus the subsample of developers with and without physical address**

	(1) Overall	(2) With developer address	(3) Without developer address
COPPA	0.358	0.354	0.371
LOCATION	0.212	0.177	0.344
DfF: Less than 5	0.128	0.153	0.036
DfF: Between 6-8	0.131	0.153	0.051
DfF: 9 and more	0.152	0.162	0.118
Log number of reviews	5.318	5.599	4.275
User rating	3.820	3.864	3.660
Top ranking ratio	0.361	0.380	0.290
Free app	0.414	0.283	0.902
Freemium	0.318	0.386	0.065
Price	0.799	0.091	0.989
Users interact	0.048	0.045	0.057
Contains ad	0.578	0.542	0.714
DfF	0.722	0.792	0.460
Without developer address	0.212	-	-
OECD	0.340	0.432	-
No OECD	0.171	0.217	-
United States	0.224	0.284	-
China	0.053	0.068	-
Member of the UE	0.264	0.335	-
Recognized by the EU	0.060	0.076	-
Independent authority	0.038	0.048	-
With legislation	0.123	0.157	-
High income	0.378	0.480	-
Upper middle income	0.055	0.069	-
Low and middle income	0.078	0.098	-
Developer: Downloads	13.033	13.638	10.789
Developer: User ratings	3.560	3.662	3.181
Developer: No User ratings	0.060	0.052	0.089
Developer missing	0.088	0.074	0.137
Observations	92746	73055	19691

Notes: Descriptive statistics of the full sample versus subsamples

Table 21: **Breakdown statistics : Subsample of developers with and without characteristics**

	(1) Overall	(2) With developer characteristics	(3) Without developer characteristics
COPPA	0.358	0.361	0.327
Users' location data	0.212	0.196	0.381
DfF: Less than 5	0.128	0.134	0.067
DfF: Between 6-8	0.131	0.134	0.105
DfF: 9 and more	0.152	0.149	0.186
Log number of reviews	5.318	5.365	4.829
User rating	3.820	3.834	3.675
Top ranking ratio	0.361	0.363	0.338
Price	0.799	0.815	0.621
Freemium	0.318	0.311	0.391
Users interact	0.048	0.047	0.056
Contains ad	0.578	0.574	0.624
DfF	0.722	0.724	0.693
Without developer address	0.212	0.201	0.331
OECD	0.340	0.351	0.222
No OECD	0.171	0.166	0.219
United States	0.224	0.225	0.208
China	0.053	0.057	0.019
Member of the UE	0.264	0.268	0.223
Recognized by the EU	0.060	0.062	0.031
Independent authority	0.038	0.038	0.041
With legislation	0.123	0.123	0.130
High income	0.378	0.381	0.348
Upper middle income	0.055	0.057	0.028
Low and middle income	0.078	0.079	0.066
Developer: Downloads	13.033	14.135	1.575
Developer: User ratings	3.560	3.864	0.397
Developer: No User ratings	0.060	0.049	0.174
Developer missing	0.088	-	-
Observations	92746	84611	8135

Notes: Descriptive statistics of the full sample versus subsample