Economics Bulletin

Volume 42, Issue 2

# Third parties in the app market and economics of privacy

Grazia Cecere
*Institut Mines Telecom, Business School, LITEM*

Vincent Lefrere
*Institut Mines Telecom, Business School, LITEM*

Fabrice Le Guel
*Université Paris-Saclay, RITM*

## Abstract

In the mobile applications market, the majority of mobile apps are free to download. We focus on some special features of this market such as access to specific user data and growing importance of third parties. Many apps features and in particular advertising, business analytics are offered via third parties which are embedded into the apps. This third parties market is hidden to consumers but it offers an important value added to app functionalities. We investigate whether personal data are more likely to be collected by apps that use third parties such as firms offering social network services or access to advanced services. We find that apps that are associated with big third parties are less likely to collect personal data. The results suggest that big firms have the competences to extract value from data without collecting too much personal data.

We thank the financial support of the Carnot TSN, the chair Good in Tech and the DAPCODS/IOTics ANR project (ANR-16-CE25-0015).
**Citation:** Grazia Cecere and Vincent Lefrere and Fabrice Le Guel, (2022) "Third parties in the app market and economics of privacy", *Economics Bulletin*, Volume 42, Issue 2, pages 1040-1049
**Contact:** Grazia Cecere - grazia.cecere@imt-bs.eu, Vincent Lefrere - grazia.cecere@imt-bs.eu, Fabrice Le Guel - fabrice.le-guel@universite-paris-saclay.fr.
**Submitted:** November 23, 2021. **Published:** June 30, 2022.

# 1. Introduction

Mobile apps are essential to many economic activities. They are used to communicate (TikTok, Snapchat), to travel via ride-sharing (Uber, Lyft) or to listen to music (Spotify, Deezer). According to Statista, annual app market revenue will rise from $318 billion in 2020 to $526 billion in 2023. The increasing availability of app-level data represent more opportunities for business analytics data performance and innovation in the market but it challenges data competition and regulation. The mobile apps market is dominated by two platforms - Google Play Store (Google) and the App Store (Apple) which at the end of 2021 accounted respectively for 3,4 million and 2,2 million apps available for download.[1] Like other free digital goods, apps are related to various revenue streams which include freemium (also called in-app purchases or integrated purchases) and advertising (Bresnahan *et al.*, 2015). In 2020, free apps (apps that are free to download without a paywall) constituted 95.8% of total apps commercialized in the Google Play Store.[2]

We provide insights into the relationships between the third-party business-to-business (B2B) market and app developers data collection strategies. Third parties are software components developed by firms and embedded in the app code.[3] They allow app developers to outsource a part of their code to improve apps (image, payment) and increase apps value (mobile analytics, advertising). Third parties are remunerated by a share of income generated by the app in the case of ad third parties or they take a commission when there is a business transaction (ex. payment third parties). While third parties are embedded in apps neither Google Play Store nor the specific apps provide information on these third parties. Third parties can access user data without the user's awareness (Razaghpanah *et al.*, 2018).

We restrict our analysis to data collected by the Google Play Store which has the largest market share (Statcounter, 2020). The sample includes 181 different third parties embedded in 239,796 apps. An observation corresponds to a given third party. In our sample, 50.4% of the apps embed at least one-third party. We have unique data based on matching apps' privacy-related characteristics evaluated by PrivacyGrade[4] to the app data collected via Google Play Store. At the time of data collection, there were five types of third parties: Advertising, Utility, Social Network, Mobile Analytics and Payment.

To each of these third parties we associate the type of personal data collected by developers at the app level. Developers decide which type of personal data to collect via a permissions system that matches their particular apps' functionalities. Permissions give access to different types of personal data such as user's location, user's pictures or contacts. There are more than a hundred different permissions available in the Google Play Store.

We estimate whether third parties size is associated to app sensitive data collection. On the one hand, third parties could encourage apps to collect more personal data in order to gather market data or improve app's functionality. On the other hand, third parties might use their market data to improve business analytics limiting personal data collection. Our results suggest that

[1] https://fr.statista.com/statistiques/571454/nombre-d-applications-disponibles-sur-les-principaux-app-stores Last accessed February 2022.
[2] https://www.appbrain.com/stats Last accessed February 2020.
[3] https://www.appbrain.com/stats/libraries/tag/ui-component/android-widgets-and-ui-libraries Last accessed February 2020.
[4] http://www.privacygrade.org Last accessed February 2020.

the biggest third parties (those used by the largest numbers of apps) are less likely to be associated with apps collecting sensitive data.

Our paper contributes to three streams of literature. First, it contributes to work on the economics of mobile apps (Ghose and Han, 2014; Li *et al.*, 2016; Comino *et al.*, 2019; Cecere *et al.*, 2020a). While there is a large body of work on the economics of mobile apps, few studies concentrate on third parties. Third parties offer a wide variety of services: app analytics, advertising, image management (Viennot *et al.*, 2014). Previous work highlights that third parties are more likely to be associated with data collection (Kesler *et al.*, 2018). Wang *et al.* (2015) show that on average third parties represent more than 60% of Android app code. Our research contributes by identifying whether third parties are associated to the collection of user data. Second, we contribute to the literature on the economics of privacy (Acquisti *et al.*, 2016). Personal data can be used to improve advertising and they can be used also to offer personalized services or to predict users behavior (Tucker, 2018). Our research adds to this literature by highlighting the role of personal data in the app market. Third, it contributes to the literature on the economics of data-driven platforms. On the one hand, the literature provides empirical evidence based on field experiments and machine learning analysis for diminishing returns to scale from additional data (Bajari *et al.*, 2019; Claussen *et al.*, 2021; Cecere *et al.*, 2020b). On the other hand, Schäfer and Sapi (2020) provide contrary empirical evidence that additional data is important and can potentially improve the quality of service and prediction of search result quality.

## 2. Third Parties Characteristics

PrivacyGrade data allow us to identify the third parties involved in each app. Our sample includes 181 third parties. Our initial dataset counts 475,787 apps and only 239,796 apps have at least one third parties. PrivacyGrade provided a categorization for five groups of third parties: Advertising, Utility, Social Networking, Mobile Analytics and Payment. Table I presents the percentage of apps using each third party category. Apps can use several third parties simultaneously. Appendix Figure 3 is an example of the third parties embedded in an app. Table IV in Appendix presents the distribution of third parties by app categories.

The variable *Advertising* indicates the advertising third parties. These are the largest group of third parties in our sample with 79 different entities. This group includes big companies such as AdMob, one of the world’s largest mobile advertising platforms which is owned by Google. The ad network transfers between 50% to 80% of the revenue generated, directly to developers. It is largely used by apps in the game category (see Table IV).

The variable *Utility* indicates utility third parties. They help developers to add functions that they did not develop. Our sample includes 72 different utility third parties including Amazon, Unity3d, and Nostra13. These third parties are largely used by the “Game all” and “Education” categories. Amazon connects mobile apps using Amazon Web Services. Unity3d is a development engine which creates interactive 3D content.

The variable *Social Networking* indicates social network third parties. This variable includes third parties such as Facebook and Twitter linking the app to social networking companies to allow consolidation of user profiles. This category includes 10 companies. They are mainly used by apps in the “Game all” and “Lifestyle” categories.

The variable *Mobile Analytics* indicates mobile analytics third parties. They are used to collect and analyze app usage and are used by 7.8% of the apps in our sample. They offer enhanced business analytics services to developers or investors. This category includes 12 companies including Yahoo owned by Flurry. This group of third parties is used largely by the "Games all" and "News and Magazines" categories. Flurry examines user data to offer business analytics services to app developers. Another company, Comscore provides independent data, metrics, products, and services to customers in the media, advertising, and marketing sectors.

The variable *Payment* includes eight different third parties which allow payment through the app. These third parties are largely used by apps in the "Lifestyle" and "Business" categories. PayPal is the most frequent third party in this group.

**Table I: Breakdown of Statistics on the Third Parties Presented in our Sample**

| Category of Third Parties | Number of Apps | Number of different Third Parties |
|---|---|---|
| Advertising | 153,978 | 79 |
| Utility | 89,153 | 72 |
| Social Networking | 65,046 | 10 |
| Mobile Analytics | 37,115 | 12 |
| Payment | 17,051 | 8 |
| Observations | | 181 |

*Notes*: This table provides summary statistics for different categories of third parties classified by PrivacyGrade. Apps can have multiple third parties.

## 3. Empirical Strategy

We provide graphical evidence on third party market concentration which might affect data collection. Figure 1 depicts the distribution of the 20% most frequent third parties in our sample. We observe that the market is skewed; AdMob[5] is the most frequently used third party followed by Facebook and Flurry, the respective leaders in the *Advertising*, *Social Networking*, and *Mobile Analytics* categories.

[5] AdMob is Google's advertising third party. The company was created in 2006 and was bought by Google in 2009 for US $750 million. More than a million apps use AdMob, resulting in payments of US $1 billion to developers since 2012.

**Figure 1: Distribution of the Most Largest Third Parties**

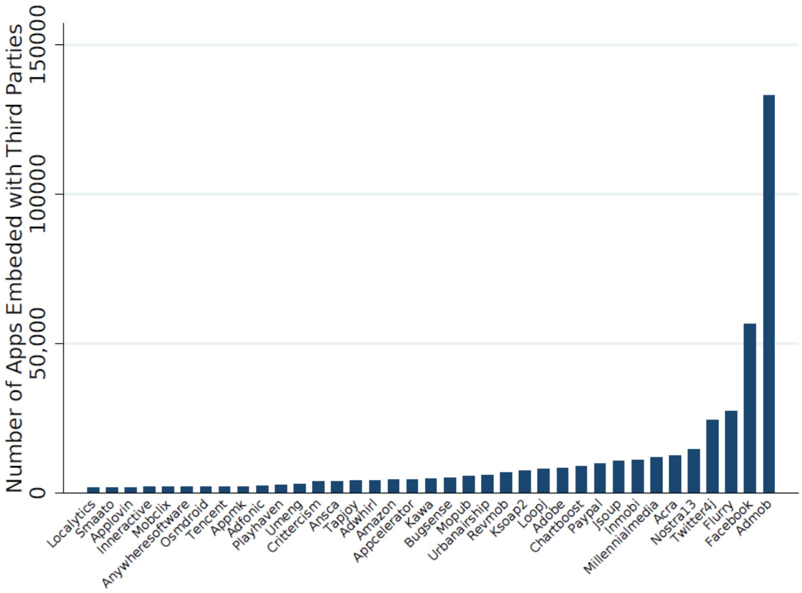


Table II reports the descriptive statistics for our key variables. The main outcome is the variable *Mean Nb Sensitive Data* which measures the average number of sensitive data permissions (used by apps) embedded in a given third party. In Table II, *Share Sensitive Data* is the percentage of apps collecting at least one piece of sensitive data embedded into a given third party. The personal data measure includes location data and the unique user identifier number (see Table V in Appendix). We investigate whether third parties size is correlated with apps collecting location or/and unique user identifier. *Share Location* is the percentage of apps collecting at least one piece of location data embedded into a given third party. *Share IMEI* is the percentage of apps collecting the permission *Read Phone Status and Identity* embedded into given third party.

*Log Num of Apps* measures the size of the third parties using the log of the number of apps embedded in each third party.

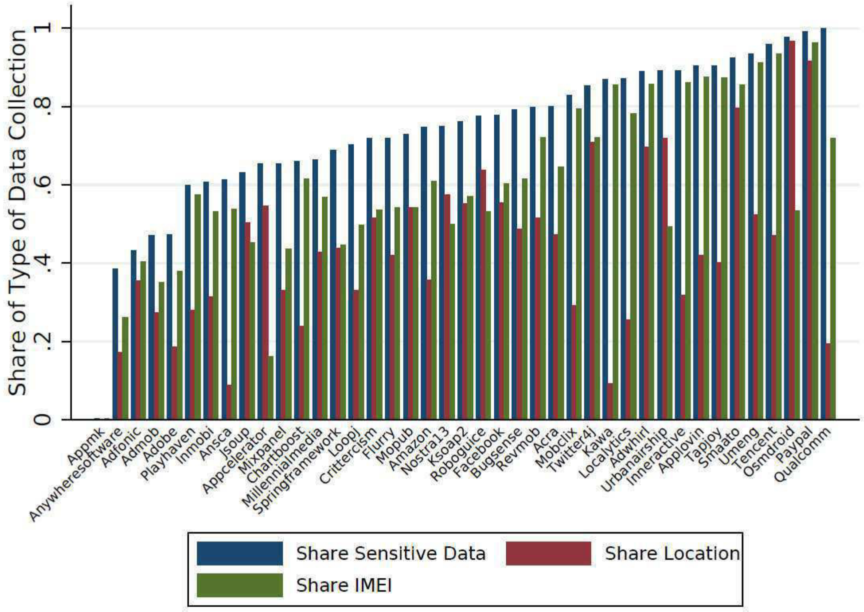
**Table II: Descriptive Statistics at the Third Parties Level**

| | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|
| Mean Nb Sensitive Data | 1.978 | 0.83 | 0 | 5.1 |
| Share Sensitive Data | 0.822 | 0.19 | 0 | 1.0 |
| Share Location | 0.554 | 0.27 | 0 | 1.0 |
| Share IMEI | 0.694 | 0.26 | 0 | 1.0 |
| Log Num of Apps | 5.587 | 2.37 | 0 | 11.8 |
| Advertising | 0.436 | 0.50 | 0 | 1.0 |
| Utility | 0.398 | 0.49 | 0 | 1.0 |
| Social Networking | 0.055 | 0.23 | 0 | 1.0 |
| Mobile Analytics | 0.066 | 0.25 | 0 | 1.0 |
| Payment | 0.044 | 0.21 | 0 | 1.0 |
| Observations | 181 | | | |

*Notes:* This table presents the descriptive statistics for the sample at the third party level.

Figure 2 depicts the distribution of the types of data collected by third parties. We observe that apps using Qualcomm, PayPal, Osmdroid, Tencent third parties are likely to collect personal data and especially unique user identifiers (or IMEI numbers).

**Figure 2: Average Sensitive Data by Third Parties**



## 4. Main Estimates

We study the correlation between third party size (measured as the number of apps linked to each third party) on app data collection. *Mean Nb Sensitive Data$_j$* is the main dependent variable where *j* is a given third party:

$$Mean\ Nb\ Sensitive\ Data_j = \beta_0 + \beta_1 Log\ Num\ of\ Apps_j + \beta_2 X_j + \epsilon_j \qquad (1)$$

*Log Num of Apps* represents the log of the number of apps associated to third party *j*. This variable measures third party size (concentration). *X* is a vector of the third party categories, the reference group is *Social Networking* (Table I). All estimates are computed with robust standard errors. Table III presents the OLS estimates. Column (1) presents the main equation. We also estimate the main specification using alternative dependent variables to investigate the type of data available to third parties. These estimates are presented in columns (2)-(4). The main variable of interest *Log Num of Apps* is significant and negative suggesting that bigger third parties are less likely to be embedded in apps collecting data. Compared to social networking third party, Utility third party is less likely to be embedded in apps collecting sensitive data. This might be because large third parties have better technical competences to exploit personal data which reduces data collection. Another possibility is that companies may benefit from having large quantities of data but with diminishing return to scale (Bajari *et al.*, 2019; Claussen *et al.*, 2021). These findings are in line with Cecere *et al.* (2020b) work which shows that in the child app market big developers collect less sensitive data on average. The pattern that large third parties are less likely to be associated with apps that collect sensitive data is replicated across the estimates.

**Table III: OLS Estimates**

| | (1)<br>Mean<br>Nb Sensitive Data | (2)<br>Share<br>Sensitive Data | (3)<br>Share<br>Location | (4)<br>Share<br>IMEI |
|---|---|---|---|---|
| Log Num of Apps | -0.079*** | -0.022*** | -0.033*** | -0.026*** |
| | (0.027) | (0.006) | (0.010) | (0.008) |
| Advertising | -0.360** | -0.045 | -0.050 | -0.000 |
| | (0.182) | (0.028) | (0.061) | (0.040) |
| Mobile Analytics | -0.636*** | -0.068 | -0.115* | -0.090 |
| | (0.200) | (0.043) | (0.069) | (0.063) |
| Utility | -0.781*** | -0.173*** | -0.222*** | -0.291*** |
| | (0.179) | (0.032) | (0.061) | (0.042) |
| Payment | -0.193 | 0.015 | -0.119 | 0.025 |
| | (0.485) | (0.037) | (0.148) | (0.068) |
| Constant | 2.936*** | 1.036*** | 0.862*** | 0.959*** |
| | (0.239) | (0.046) | (0.086) | (0.064) |
| Observations | 181 | 181 | 181 | 181 |

*Notes:* OLS regression estimates. Dependent variable as described in column headers. Observations are at the third party level. The reference category is *Social Networking*. Robust Standard errors in parentheses. Significance levels: $*p < .10$, $**p < .05$, $***p < .01$

# 5. Conclusion

While third parties are essential to enable certain app functionalities, little is known about the structure of the third party market. We highlight that the presence of third parties is important for the provision of enhanced services and features which promote the creation of new innovative companies. We show that large third parties (those used by many apps) are less likely to be associated with apps collecting sensitive data. In this market, data concentration seems to reduce sensitive data collection. Our paper underlines that the link between third parties and sensitive data is likely to be negative.

Our results contribute to the privacy regulation debate. On the one hand, unlimited access by third parties to users' personal data raises privacy concerns. On the other hand, third parties allow feedback on app functioning alerting developers to bugs and consumer usage. In addition to supplying basic information on functionality, third parties provide results related to content, improved user experience, app quality, and reduce developer costs.

This lack of knowledge about the ultimate use of sensitive data by third parties is crucial especially since the user is often unaware that a third party is embedded in the app. Since the regulatory authorities need to reconcile competition with user privacy, the evidence we provide on data concentration and the link to personal data collection should be informative.

# References

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

Bajari, P., Chernozhukov, V., Hortaçsu, A. and Suzuki, J. (2019) "The Impact of Big Data on Firm Performance: An Empirical Investigation" *AEA Papers and Proceedings* **109**, 33–37.

Bresnahan, T., Davis, J., Jaconette, T. and Yin, P.-L. (2015) "Mobile Applications, the Economics of." in *The New Palgrave Dictionary of Economics*, Springer, 1-8.

Cecere, G., Le Guel, F. and Lefrere, V. (2020a) "Economics of Free Mobile Apps: Personal Data and Third Parties" Working Paper.

Cecere, G., Le Guel, F., Lefrere, V., Tucker, C. and Yin, P.-L. (2020b) "Privacy, Data and Competition: The Case of Apps For Young Children" Working paper.

Claussen, J., Peukert, C. and Sen, A. (2021) "The Editor vs. the Algorithm: Returns to Data and Externalities in Online News" Working paper.

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.

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

Kesler, R., Kummer, M. and Schulte, P. (2018) "Mobile Applications and Access to Private Data: The supply Side of the Android Ecosystem" ZEW-Center for European Economic Research Discussion Paper.

Li, X., Bresnahan, T. and Yin, P.-L. (2016) "Paying Incumbents and Customers to Enter an Industry: Buying downloads" Working paper.

Razaghpanah, A., Nithyanand, R., Vallina-Rodriguez, N., Sundaresan, S., Allman, M., Kreibich, C. and Gill, P. (2018) "Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem" in *NDSS*.

Schäfer, M. and Sapi, G. (2020) "Learning from Data and Network Effects: The Example of Internet Search" DIW Berlin Discussion Paper.

Statcounter (2020) "Mobile Operating System Market Share Worldwide" Retrievable at https://gs.statcounter.com/os-market-share/mobile/worldwide.

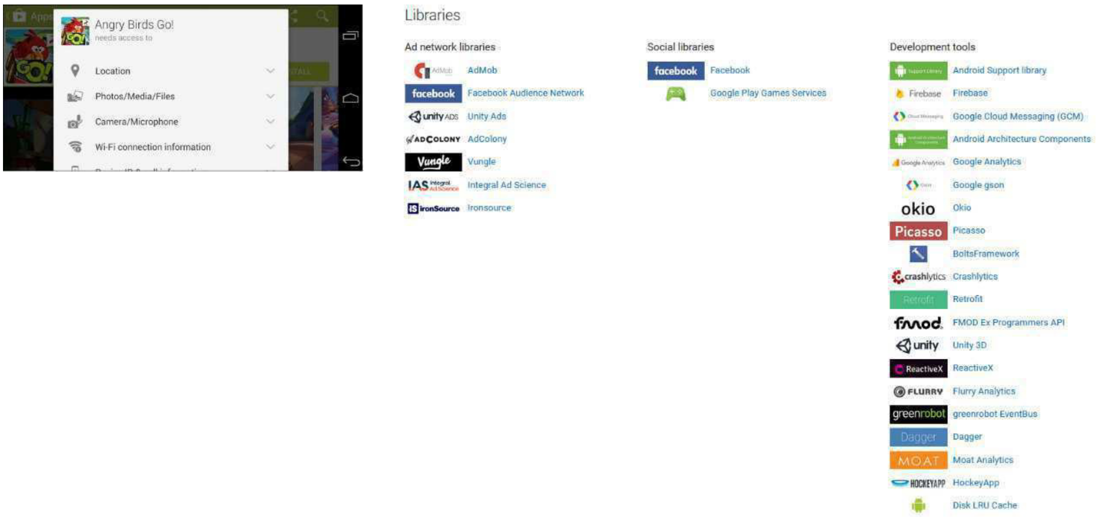
Tucker, C. (2018) "Privacy, Algorithms, and Artificial Intelligence. In The Economics of Artificial Intelligence: An Agenda" University of Chicago Press.

Viennot, N., Garcia, E. and Nieh, J. (2014) "A Measurement Study of Google Play" in *ACM SIGMETRICS Performance Evaluation Review,* ACM, 221–233.

Wang, H., Guo, Y., Ma, Z. and Chen, X. (2015) "Wukong: A scalable and Accurate Two-Phase Approach to Android App Clone Detection" in *Proceedings of the 2015 International Symposium on Software Testing and Analysis*, ACM, 71–82.

# Appendix

## Figure 3: Example of Third Parties and Permissions Embedded Into an App



Angry Birds Go!
needs access to

- Location
- Photos/Media/Files
- Camera/Microphone
- Wi-Fi connection information

Libraries

Ad network libraries

- AdMob
- Facebook Audience Network
- Unity Ads
- AdColony
- Vungle
- Integral Ad Science
- Ironsource

Social libraries

- Facebook
- Google Play Games Services

Development tools

- Android Support library
- Firebase
- Google Cloud Messaging (GCM)
- Android Architecture Components
- Google Analytics
- Google gson
- Okio
- Picasso
- BoltsFramework
- Crashlytics
- Retrofit
- FMOD Ex Programmers API
- Unity 3D
- ReactiveX
- Flurry Analytics
- greenrobot EventBus
- Dagger
- Moat Analytics
- HockeyApp
- Disk LRU Cache

Source: AppBrain.com, 2020

## Table IV: Breakdown Statistics of the Third Parties Presented in Our Sample

| | Third Parties Categories | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| App Categories | Advertising | Utility | Social Networking | Mobile Analytics | Payment |
| | % | % | % | % | % |
| Books and Reference | 5.96 | 5.18 | 2.50 | 2.36 | 1.33 |
| Business | 3.49 | 7.86 | 9.10 | 5.17 | 22.06 |
| Comics | 0.37 | 0.28 | 0.20 | 0.23 | 0.11 |
| Communication | 1.46 | 2.45 | 1.27 | 1.15 | 0.85 |
| Education | 7.60 | 8.17 | 6.73 | 6.65 | 7.18 |
| Entertainment | 8.58 | 7.19 | 7.60 | 5.81 | 7.70 |
| Finance | 1.59 | 3.03 | 1.28 | 1.69 | 1.67 |
| Games | 27.54 | 16.32 | 22.19 | 30.16 | 6.49 |
| Health and Fitness | 2.69 | 2.80 | 3.66 | 2.83 | 4.47 |
| Libraries and Demo | 0.12 | 0.22 | 0.07 | 0.05 | 0.11 |
| Lifestyle | 6.19 | 7.64 | 10.47 | 5.78 | 15.38 |
| Media and Video | 1.21 | 1.33 | 0.85 | 1.00 | 0.35 |
| Medical | 0.90 | 1.60 | 1.45 | 1.42 | 2.02 |
| Music and Audio | 4.19 | 3.66 | 4.59 | 4.31 | 3.96 |
| News and Magazines | 3.88 | 5.47 | 4.81 | 8.22 | 5.68 |
| Personalization | 2.95 | 1.17 | 0.53 | 2.62 | 0.25 |
| Photography | 1.37 | 1.42 | 1.41 | 1.24 | 0.46 |
| Productivity | 2.21 | 2.92 | 1.52 | 1.91 | 1.64 |
| Shopping | 1.13 | 2.00 | 2.30 | 1.63 | 2.65 |
| Social | 1.81 | 2.67 | 3.58 | 2.17 | 1.82 |
| Sports | 2.66 | 3.62 | 4.04 | 3.07 | 5.41 |
| Tools | 6.87 | 4.47 | 1.94 | 3.28 | 1.08 |
| Transportation | 1.17 | 1.73 | 0.96 | 1.43 | 1.05 |
| Travel and Local | 3.53 | 6.35 | 6.74 | 5.01 | 6.02 |
| Weather | 0.53 | 0.47 | 0.23 | 0.79 | 0.27 |

*Notes*: This table provides the distribution of the different categories of third parties classified by PrivacyGrade inside the Google Play Store categories. Column (1) presents the distribution of apps using advertising third parties. Column (2) the distribution of apps using utility third parties. Column (3) presents the distribution of apps using social networking third parties. Column (4) presents the distribution of apps using mobile analytics third parties. Column (5) presents descriptive statistics for apps using payment third parties.

## Table V: List of Permissions Used to Construct the Dependent Variables

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Mean | Share | Share | Share |
| | Nb Sensitive Data | Sensitive Data | Location | IMEI |
| Approximate Location (Networkbased) | ✓ | ✓ | ✓ | |
| Precise Location (GPS and Networkbased) | ✓ | ✓ | ✓ | |
| Mock Location | ✓ | ✓ | ✓ | |
| Record Audio | ✓ | ✓ | | |
| Take Pictures and Videos | ✓ | ✓ | | |
| Read Phone Status and Identity (IMEI) | ✓ | ✓ | | ✓ |

*Notes:* List of Permissions used to construct the dependent variables. Column (1) presents the count variable *Mean Nb Sensitive Data*. Column (2) presents the dichotomous variable *Share Sensitive Data* equals one if at least one permission was chosen by the developer. Column (3) presents the dichotomous variable *Share Location* equals one if at least Approximate Location or Precise Location or Mock Location permissions was chosen by the developer. Column (4) presents the dichotomous variable *Share IMEI* equals one if Read Phone Status and Identity (IMEI) was chosen by the developer.