Impact of App Privacy Label Disclosure on Demand: An Empirical Analysis

Rajiv Garg[a], *Rahul Telang*[b]

[a] Goizueta Business School, Emory University

[b] Heinz College, Carnegie Mellon University

Abstract

Personal data privacy is becoming an increasing concern, especially amongst the mobile app users. With increasing push from the policy makers, mobile platforms like Google Android and Apple iOS mandated app developers to disclose user data collected by their apps. Since role of privacy disclosure on consumer demand remains a big unknown, we use privacy label disclosure policy to evaluate the role of these labels on the demand of top downloaded apps on the iPhone. We find that app developers are strategic in updates when they are collecting more intrusive user information, and disclosure of privacy label to collect more sensitive information reduces app demand. We further find negative effect on demand when large number of competing apps start disclosing privacy labels. Additionally, from within app category analysis, we find that it is better for apps collecting sensitive data (e.g., social media app) to disclose early versus delay the disclosure.

Introduction

There is a widespread belief that "we are fighting a losing battle with privacy". Whether on the Internet or otherwise, more and more data about us is being generated faster from more devices, and we can't keep up. It's a losing game both for individuals and our legal system. There have been many well publicized privacy violations[1] and data breaches[2] and disputes between firms and FTC (Federal Trade Commission). With privacy as a focus, GDPR (General Data protection regulation) was introduced in Europe in 2019. USA has also followed up with various regulation like CCPA (California Consumer privacy Act).

Clearly the balance between privacy and commerce is complex that we cannot agree on a consensus (Auxier et al. 2019). These above noted regulations are geared towards changing firm behavior so both user privacy and firm's commercial interests are maintained. Firms realize the fallout and have taken many steps in improving user privacy. For example, banning use to third party cookies[3] has been a move by firms like Mozilla, Apple, and Google. The goal is to not collect user data to serve them personalized ads as it causes user discomfort about their privacy. Recently, Apple has instituted mandatory privacy labels for all the apps on its App store. Apple has a large market

[1] https://www.nbcnews.com/tech/social-media/timeline-facebook-s-privacy-issues-its-responses-n859651

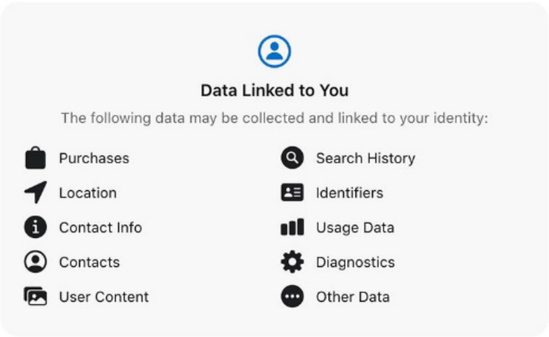
[2] https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement

[3] Which is used heavily in personalized advertisement

share and Appstore has close to 1.8 million apps for download[4] that generated almost $64B in revenue.[5] Given the popularity and share of Appstore, privacy labeling has a huge impact on App developers as they must indicate privacy labels associated with their apps on Appstore. The goal is to educate users of their data that is collected by the App. Since there are so many choices, they can choose to go with the App that is violating their privacy the least. Informing user about firm practices about data protection is a widely used policy - like data breach disclosure laws (Romanosky et al. 2011). The goal is to inform the users about their data firms will be sharing so they can make informed choices; this is akin to food labeling (Golan et al. 2001).

In many digital products, lack of information is a significant concern. Privacy is complex because it is impossible for users to know how their privacy is invaded when they use an App or any software. Without enough details, they cannot make a right choice of choosing a product even if they want more security and privacy. The goal of such disclosures is to provide information to users in easy-to-understand labels so they can make optimal choices.

With high penetration of mobile devices and applications, data is tracked by both software and hardware providers. Not surprisingly, data tracking generates concern for privacy amongst consumer (Barkhuus and Dey 2003). A recent study (Boyles et al. 2012) found that 54% of app users decide not to install a mobile app when they are made aware of their private data being collected, and 30% of app users uninstalled the app. In the light of this, Apple announced that all the apps on their Appstore will have to display privacy labels after December 14, 2020, so that users can learn about the types of data an app collects (types of data -not liked to a user, linked to a user, used to track a user[6]). The purpose of the label is to help customers understand what data is being collected by an app and how that data is used so they can make informed decision. A sample label is presented in figure 1.



Data Linked to You

The following data may be collected and linked to your identity:

- Purchases
- Location
- Contact Info
- Contacts
- User Content
- Search History
- Identifiers
- Usage Data
- Diagnostics
- Other Data

Figure 1: Privacy Label on App Page in App Store

Since many Apps rely on user data for revenue, there is significant concern that use of such data is infringing on user privacy. While some data is useful to provide personalized features, much of the data is used to provide targeted advertisements. The question is whether providing information changes user demand. In this research, we investigate the impact of these privacy labels on the App demand.

[4] https://www.businessofapps.com/data/app-statistics/

[5] https://www.cnbc.com/2021/01/08/apples-app-store-had-gross-sales-around-64-billion-in-2020.html

[6] https://developer.apple.com/news/?id=3wann9gh

User's privacy concerns on online platforms have been studied extensively in the prior literature (Bélanger and Crossler 2011) and show that consumers often lack information to make privacy-sensitive decisions (Acquisti and Grossklags 2005). The surveys typically suggest that users care about privacy and having more information will lead to them making more informed decision. Thus, one would intuitively think that if an App discloses that it is collecting more privacy invasive data, the users will punish the App and hence reduce the demand. Thus, privacy label disclosure may influence consumers' decision to (or not to) adopt an app. It has been widely reported in many contexts that giving more information to users lead to changes in user behavior (Adjerid et al. 2018; Kim et al. 2008). However, it is not clear that such nudges will impact demand for Apps. These are many factors including information overload, inability of users to evaluate alternatives, not trusting that labels provide appropriate information about privacy intrusion, that may lead to these labels being ineffective.

The app developer's decision to display various labels and speed with which they display them will also be a function of competition. As a result, an app developer may choose to disclose their privacy labels if they can disclose only a few labels to gain advantages from in the mobile app market. Thus, in this extended abstract, we provide the results for the role of number of data types in privacy labels on the rank of the apps and the relationship between the app update duration to disclose labels and the number of label attributes disclosed.

Data & Insights

To empirically understand the role of privacy label disclosure on demand, we collected data from Nov 1, 2020 to June 30, 2021, on iOS AppStore, an App's rank (both overall and within a category), the content of its label, and when they are displayed. We collect top 100 Apps for each of the 22 categories listed in App store. This gives us most of the important Apps sold on AppStore in each category. Needless to say, this data collection is a non-trivial task, and it is still going on. There are 3 types of labels listed in privacy labels for an app - "Data Used to Track You" (TrackYou)[7], "Data Linked to You" (LinkYou), and "Data Not Linked to You" (NotYou). Table 1 below shows the summary statistics by category of these apps and their labels. As seen below, in "photo & video" category there are 0.48 data types for "TrackYou", 2.96 data types for "LinkYou", and 0.64 data types for "NotYou". We see that "food and drink" apps collect most data types, and "Business" and "Navigation" apps don't collect any tracking data. This is somewhat intuitive because of the extent of personalization each app in those categories may provide.

Table 1: Summary Statistics by App Category
(Apps that were present in ranked list before and after labels were added)

Category	Price	Rank (overall)	Avg (rating)	LinkYou (count)	NotYou (count)	TrackYou (count)
Reference	2.42 (1.16)	17.46 (18.24)	4.71 (0.32)	0.89 (2.27)	2.18 (1.31)	0.04 (0.29)
Productivity	1.78 (2.7)	27.74 (26.11)	4.42 (0.6)	2.71 (3.85)	1.16 (1.43)	0.25 (0.44)
Sports	1.53 (1.22)	67.21 (19.87)	4.4 (0.71)	1.45 (2.88)	1.02 (1.39)	0.61 (1.61)
Social Networking	0.18 (0.71)	38.31 (26.08)	4.27 (0.83)	5.23 (4.4)	0.47 (0.74)	0.85 (1.39)
Utilities	5.76 (8.99)	42.6 (27.7)	3.94 (0.57)	1.57 (3.73)	0.28 (0.7)	0.02 (0.21)

[7] It is more privacy intensive

Travel	0.32 (1.48)	55.71 (15.7)	4.74 (0.15)	8 (4.35)	0.5 (0.74)	1.87 (2.67)
Business	1.03 (2.27)	23.91 (9.13)	4.71 (0.13)	3.67 (3.43)	0.22 (0.52)	0 (0)
Weather	4.62 (2.26)	85.13 (11.49)	4.65 (0.12)	0.33 (1.56)	0.02 (0.26)	0.24 (1.16)
Photo & Video	3.33 (4.24)	34.32 (23.41)	4.26 (0.64)	2.96 (4.26)	0.64 (1.25)	0.48 (0.85)
News	0.01 (0.15)	53.69 (21.94)	4.66 (0.35)	0.41 (1.66)	4.73 (3.53)	1.62 (1.22)
Music	3.08 (5.08)	43.59 (26.38)	4.45 (0.67)	3.21 (3.81)	1.14 (1.75)	1.66 (1.8)
Navigation	0.48 (0.92)	39.15 (19.77)	4.61 (0.29)	4.29 (4.88)	0.23 (0.45)	0 (0)
Health & Fitness	2.89 (1.9)	53.07 (19.88)	4.6 (0.32)	2.3 (2.24)	1.1 (0.95)	0.96 (1.86)
Lifestyle	0.06 (0.33)	41.47 (24.08)	4.46 (0.47)	5.55 (4.11)	0.03 (0.31)	0.57 (1.18)
Finance	0.01 (0.13)	71.37 (22.9)	4.54 (0.36)	7.8 (4.92)	0.56 (0.83)	0.47 (1.05)
Games	1.61 (1.59)	41.82 (24.95)	4.29 (0.63)	0.48 (1.94)	1.73 (1.53)	0.6 (1.11)
Education	9.36 (10.7)	59.35 (28.86)	4.09 (0.99)	1.54 (2.52)	1.09 (2.23)	0.35 (1.01)
Entertainment	0.02 (0.21)	44.81 (33.38)	4.68 (0.29)	5.7 (4.32)	0.57 (0.89)	1.49 (1.5)
Medical	5.99 (5.92)	52.2 (17.92)	3.73 (0.41)	2.55 (2.44)	0.79 (0.71)	1.29 (1.49)
Food & Drink	0.33 (1.24)	68.65 (24.03)	4.56 (0.28)	6.83 (4.75)	0.14 (0.63)	3.16 (2.7)
Shopping	0.01 (0.13)	69.5 (22.98)	4.79 (0.1)	6.67 (4.85)	0.77 (1.01)	1.65 (2.32)
Overall	4.72 (1.07)	42.25 (13.05)	3.94 (0.17)	0.04 (0.35)	0.11 (0.54)	0.06 (0.29)

In addition, from table 2 below, we see that "top free apps" collect more data from users compared to the "top paid apps." This is intuitive because free apps need additional data to monetize using targeted advertising. Top paid apps tend to provide some niche features/functionality in the app store that allows them to command a price and thus they are more averse to collecting additional customer data to enable some functionality (e.g., purchases). Top revenue grossing apps have characteristics like free apps because that list is generally dominated by free apps that are selling premium services within the app.

Table 2: Average Number of Data Types in Each Label

	TOP FREE APPS		TOP PAID APPS		TOP GROSSING APPS	
	obs	mean	obs	mean	obs	mean
SUM (Data collected to track you – TrackYou)	73,671	3.109	4,485	2.234	75,359	3.054
SUM (Data linked to you – LinkYou)	76,172	5.299	6,046	3.238	76,167	5.033
SUM (Data not linked to you – NotYou)	69,825	2.835	17,770	2.400	65,750	2.880

Further exploration of the number of labels disclosed by apps in a given category (figure 2 – app-category on x-axis), we find that, across all app categories, free apps collect more data linked to a user whereas paid apps tend to have more privacy data types classified under data not linked to user. This could be the case because free apps need user data to monetize and to target ads within the app. Paid apps tend to not create aversion in purchase of the app, top grossing apps collect data linked to a user for monetization through upselling or cross-selling.

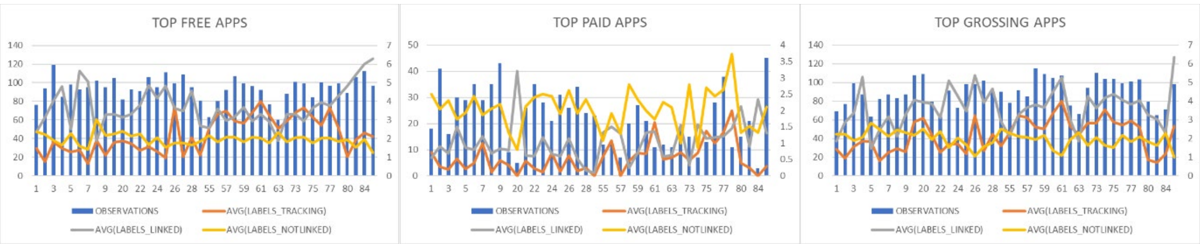


Figure 2: Average Labels Disclosed (by app-category)

Observing the frequency distribution (histogram) of the label disclosure timing by categories (figure 3), we observe that the app categories that are likely to have more invasive private data needs (e.g., shopping or social media apps) tend to update sooner – shopping apps generally need more data linked to a user to better target users across apps (Molitor et al. 2020). Similarly, apps that need less invasive data (e.g., photo/video or productivity apps) push app developers to delay updates to their apps. Overall, we do observe that the app developer's update decisions are strategic based on their app needs and app category (Mayya and Viswanathan 2021). But there is a lack of understanding if these disclosures causally affect the demand (downloads/ranks) of these apps, and how the demand is affected for various app categories and privacy label data types. The goal of this paper is to answer these questions around the role of information disclosure on demand.

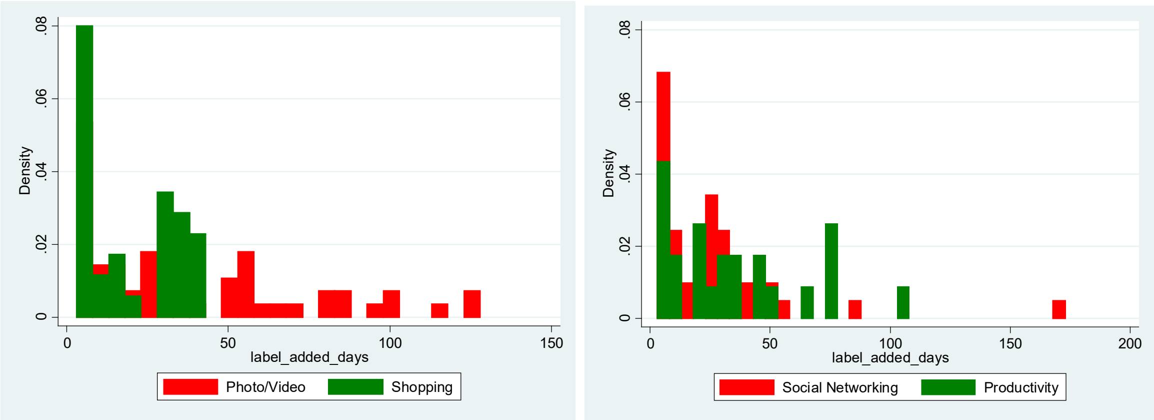


Figure 3: Privacy Label Disclosure Duration Histogram (four app categories)

To better understand the relationship between the labels and the update timing, we present the correlation matrix for data type labels and the number of days since the enactment of the policy and the actual disclosure of the labels (Appendix B - figure 5). The somewhat interesting (yet intuitive) observations is that the correlations between more intrusive privacy data types (under label "*data to track a user*") are all negative. We also observe that apps collecting "user contact information" and "user identifiers" take the longest to update and disclose that information. For privacy labels under "*data linked to the user*," two additional labels show strong negative correlation with the update duration: "user purchases" and "user content (emails, messages, photos, etc.)". Interestingly, apps that have some social component and are linking "user's contacts" or "sensitive" demographic and personal data, tend to disclose that sooner. We believe this to be the case because of the core functionality offered by those apps. Apps that are disclosing collection of data "*not linked to a user*" update those apps quickly and have a very small magnitude (positive or negative) correlation between the days to update and the privacy labels disclosed. Overall, the observations from the correlation suggest that app developers tend to update their apps later if they need to disclose the use of more invasive information (Mayya and Viswanathan 2021). Interestingly, a quick disclosure of privacy labels that are suggesting more intrusive information collection is not very intuitive. It would be possible if that app developers believe that disclosure could lead to customer trust, which could result in increased demand.

Counter to this prior intuition, when observing change in rank a week before privacy label disclosure and a week after the disclosure for each app category (figure 5), we find that average rank of apps before disclosure was slightly higher (numerically smaller) compared to the average rank post disclosure). This trend is observed for all categories in all three types of apps (top free, top paid, and top grossing). While this does provide some model-free evidence that the disclosure of privacy labels effects the demand negatively. We will build on these findings in our empirical analysis section.

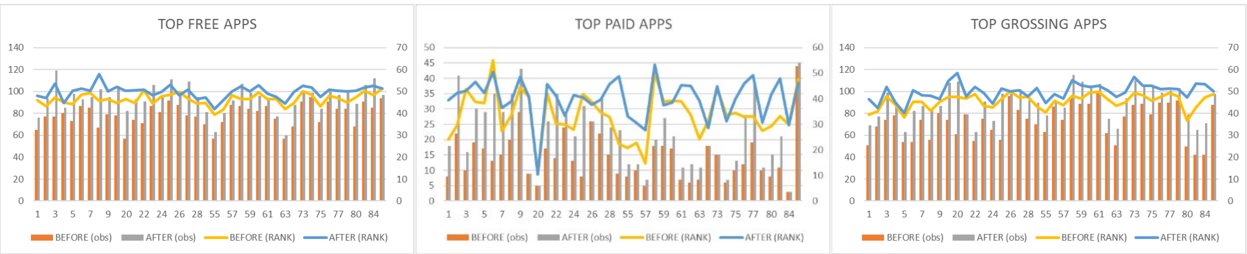


Figure 4: Average App Rank (by category) Day Before and Day After Privacy Label Disclosure

Empirical Analysis & Results

To understand the impact of disclosed privacy labels on demand, we use app (i) rank (y) as a proxy for demand (Garg and Telang 2013) at time (t) and use the three categories of privacy labels – "*Data Used to Track You*" (TrackYou), "*Data Linked to You*" (LinkYou), and "*Data Not Linked to You*" (NotYou). As seen below, we use three different empirical models where independent variables (X^{ttu}, X^{lin}, X^{not}) are: i) dummy variables identifying a data type label present in that privacy category (equation 1), ii) continuous variable measuring the number of data type labels present in the category (equation 2), and iii) dummy variable for each data type label in each privacy category (equation 3).

$$y_{it} = \beta_0 + \beta_1 X_{it}^{dtt} + \beta_2 X_{it}^{dnot} + \beta_3 X_{it}^{dlin} + \beta_4 X_{it} + \delta_i + \delta_{i\in C} + \varepsilon_{it} \tag{1}$$

$$y_{it} = \beta_0 + \beta_1 X_{it}^{ctt} + \beta_2 X_{it}^{cnot} + \beta_3 X_{it}^{clin} + \beta_4 X_{it} + \delta_i + \delta_{i\in C} + \varepsilon_{it} \tag{2}$$

$$y_{it} = \beta_0 + \beta_1 \textstyle\sum_{tt} X_{it}^{dtt} + \beta_2 \textstyle\sum_{not} X_{it}^{dnot} + \beta_3 \textstyle\sum_{lin} X_{it}^{dlin} + \beta_4 X_{it} + \delta_i + \delta_{i\in C} + \varepsilon_{it} \tag{3}$$

In the above empirical models, y_{it} is the rank of an app (i) at time (t), δ_i is the app (i) fixed effect, δ_c is the app category (c) fixed effect, Dit is the dummy variable with value "0" before label disclosure and value "1" after labels are introduced, and ε_{it} is the error term. In equation 1, X^{dttu}, X^{dnot}, and X^{dlin} are dummy variables representing the privacy label category present on the app (i) page after the update. In equation 2, X^{cttu}, X^{cnot}, and X^{clin} are count variables representing the number of data types collected by an app as listed in each of the privacy label category. In equation 3, X^{dttu}, X^{dnot}, and X^{dlin} are dummy variables representing the data type collected by an app as listed in each of the privacy label category[8]. For all empirical models, we use panel regression model for all apps that were ranked before and after the disclosure of privacy labels.

In table 3 below, we present results for equation 1 and 2 respectively for all three app types (top free, top paid, and top grossing) where each app's rank is measured within their respective category. From this table, we observe that disclosure of privacy labels ("TrackYou", "LinkYou", and "NotYou") has significant and positive effect for top free apps. Disclosure of label under "Data to Track You" (TrackYou), leads to a 0.49% drop (numerical increase) in ranking of the apps. Similarly, "Data Linked to You" leads to, on average, a 0.52% drop (numerical increase) in ranking of the apps, and disclosure of label under "Data Not Linked to You" leads to, on average, a 0.46% drop (numerical increase) in ranking of top ranked free apps. Similarly, for top grossing apps, disclosure of "Data to Track You" leads to an 0.35% drop (numerical increase) in within category ranking. Results for top paid apps are statistically insignificant.

In case of top free apps, coefficient for Xttu (count) is positive and significant suggesting that every data point that is collected to track a user the ranking declines by 0.07%. At average value of number of labels (3), the change in ranking for every additional data point collected is 0.21%. Similarly, the impact of an additional data type collected under "LinkYou" is 0.58%, and impact of an additional data type collected under "NotYou" is 0.58%. While these numbers seem small, it is important to note that when a top ranked (#1) app drops by 1 rank, the demand drops significantly because of Pareto distribution (Garg and Telang 2013). Interestingly, we find no statistical significance in coefficients for paid apps.

One limitation to these results could be possible complex interaction effects of these label types that we are not capturing in the current model. Thus, we do plan to explore potential non-linearities and multi-dimensional interaction effects in the model as we develop this paper.

Table 3: Panel Regression Model (30-days before and update update)

Ln(rank_cat)	TOP FREE APPS		TOP PAID APPS		TOP GROSSING APPS	
TrackYou (bool)	0.0048 (0.001)***		0.004 (0.004)		0.0035 (0.001)***	
LinkYou (bool)	0.0052 (0.001)***		0.0032 (0.004)		-0.001 (0.001)	
NotYou (bool)	0.0046 (0.001)***		0.0009 (0.002)		0.0005 (0.001)	
TrackYou (count)		0.0007 (0)**		0.0004 (0.002)		0.0006 (0)*
LinkYou (count)		0.0012 (0)***		0.0009 (0.001)		-0.0001 (0)
NotYou (count)		0.002 (0)***		0.0012 (0.001)		0.0005 (0)**
_cons	2.8987 (0.037)***	2.8972 (0.037)***	2.6532 (0.044)***	2.6529 (0.044)***	3.8805 (0.109)***	3.8808 (0.109)***
Category FE	yes	yes	yes	yes	yes	yes
App FE	yes	yes	yes	yes	yes	yes

[8] Please refer to appendix A for a list of all datatypes presented in each privacy label

From within category analysis (equation 3), we observe (tables 4a and 4b) that the largest drop for free apps is for the category “Games - Strategy” (0.0361) suggesting a drop in rank by 3.7% because of “TrackYou” label disclosure. Similarly, the largest drop for “LinkYou” label disclosure is for “Shopping” category (0.0223) suggesting a drop of 2.3% in ranking. The largest drop for “NotYou” label disclosure is for app category “Games – Casino” and non-game category “Utilities” seeing a drop of 2.4% and 1.5% in rankings, respectively. We also find that some app categories see an improvement in ranking possibly because of the trust created by the disclosure – for example “Sports” apps disclosing the data collected to track users. This is the key benefit of privacy information disclosure where the data collected is disclosed but customer is also notified that the data collected does not contain any identifying personal information. Furthermore, this raises the question on what data points collected within each privacy label category impact the demand. We present the preliminary results in Appendix B.

Table 4a: Regression Models by App Category – Top Free and Paid Apps

App Category	#Apps	TOP FREE			TOP PAID		
		X^{dttu}	X^{dlin}	X^{dnot}	X^{dttu}	X^{dlin}	X^{dnot}
Magazines&Newspapers	381	0.00727**	0.0153***	0.00807***		0	0.00163***
News	369	-0.00315	-0.0150***	0.00627**	0.00440	-0.00193	0.000204
Sports	356	-0.0147***	0.0154***	0.00292	-0.0205	0.00588	-0.0169*
Weather	352	-0.00744*	0.00812*	-0.0238***	0.00332	0.00135	-0.00891***
Photo&Video	308	-0.00577***	-0.00335**	-0.0127***	-0.00200	-0.00199	0.000154
Productivity	306	0.000607	0.00426***	0.00811***	0.00169	0.000171	0.0112**
Travel	291	0.00500**	-0.0178***	-0.00748***	-0.00835	-0.0193	-0.000789
Medical	286	0.00679	-0.0166***	0.0110***	-0.0165	-0.00905	-0.00516
Business	280	-0.00143	0.0133***	0.00259	0.0136	0.0272	-0.0244*
Food&Drink	262	-0.00177	-0.00104	0.00281	0.111***	-0.0243**	-0.00784
Utilities	253	-0.00369	0.0114***	0.0145***	-0.0487***	0.0389***	0.00635
Social Networking	243	0.00194	0.00103	0.00215	0.0116	-0.00435	-0.00344
Reference	229	0.00398*	0.00726***	0.00242	-0.0540***	0.0535***	0.000442
Shopping	223	0.00742***	0.0223***	0.00525**		-0.00496	-0.00896
Games -Word	343	0.0130***	0.00807***	-0.00221	-0.00229	0.0153	0.00397
Games -Board	334	0.0177***	0.00686	0.0174***	0.00758	0.0206*	0.0138***
Games -Casual	331	0.0133***	-0.00323	0.00986***	0.00428	0.000588	-0.00137
Games -Trivia	324	0.0189***	0.0130***	0.00460**	0.0121*	0.0130	-0.00899***
Games -Family	316	-0.0116**	0.0262***	0.0188***	-0.0109	0.0463***	0.00444
Games -Music	315	-0.0227***	0.0207***	0.00885***	0.0106**	0.00490	-0.00555***
Games -Sports	314	-0.000894	-0.0145***	0.0176***	-0.00943***	5.86e-05	0.00831***
Games -Action	300	-0.00604*	0.00285	-0.0134***	-0.0140	0.00874	0.00804
Games -Puzzle	291	-0.0165***	0.0142***	0.0130***	-0.0146*	-0.00167	0.00396
Games -Dice	279	-0.000548	-0.00344	0.00207	-0.0223*	-0.00474	0.0232***

Games -Educational	277	0.00734***	7.11e-05	-0.0141***	0.00380	-0.0159***	0.00988***
Games -Strategy	276	0.0361***	-0.0102***	0.00324	0.00243	0.0188	-0.0106*
Games -Card	268	0.000266	0.00552**	0.00896***	-0.0172***	-0.00326	0.0108***
Games -Casino	260	0.00385	-0.00265	0.0234***		-0.00781	0.0123
Games -RolePlaying	230	0.00627**	-0.0147***	-0.00194	-0.00122	-0.00171	-0.000357
Games -Adventure	189	0.00428	0.000588	-0.00137	0.00428	0.000588	-0.00137
Games -Simulation	181	0.000190	0.00807***	-0.00574*	0.000903	-0.000903	0.00466***
Games -Racing	126	-0.0146*	-0.00167	0.00396	-0.0146*	-0.00167	0.00396

Table 5b: Regression Models by App Category - Top Grossing Apps

App Category	#Apps	TOP GROSSING		
		X^{dttu}	X^{dlin}	X^{dnot}
Magazines&Newspapers	381	-0.000787	0.0259***	-8.78e-05
News	369	-0.0112***	-0.000556	-0.000749
Sports	356	-0.00122	-0.00149	0.00286
Weather	352	-0.00953**	-0.00700	0.00212
Photo&Video	308	0.00144	-0.00568***	0.00385***
Productivity	306	0.00219	-0.00215	0.00260
Travel	291	-0.0133**	0.000895	-0.00468
Medical	286	-0.000440	-0.00941**	-0.00833**
Business	280	0.0124***	0.00725***	0.00599***
Food&Drink	262	-0.0141**	-0.00620	-0.00215
Utilities	253	-0.00272	0.0138***	0.0181***
Social Networking	243	-0.00240	0.00741***	0.00664***
Reference	229	-0.00454*	0.00960***	0.00455**
Shopping	223	0.000783	0.0103	0.000583
Games -Word	343	0.00649***	0.00677***	-0.00327**
Games -Board	334	0.00692***	0.00158	0.00118
Games -Casual	331	-0.0103***	-0.00639***	-0.00914***
Games -Trivia	324	0.00297	-0.00307	0.00291
Games -Family	316	0.0171***	-0.00115	0.00247
Games -Music	315	0.00394**	-0.00124	0.000682
Games -Sports	314	-0.00812***	-0.000448	0.00256*
Games -Action	300	-0.000173	-0.00123	-0.000286
Games -Puzzle	291	0.00674**	0.00265	-0.00791***
Games -Dice	279	0.00406*	-0.00131	-0.00314*
Games -Educational	277	0.000787	-0.00184	-0.00146
Games -Strategy	276	0.00186	0.00118	0.00343***
Games -Card	268	0.000802	0.000276	0.00163**

Games -Casino	260	0.00316**	0.00792***	0.00214
Games -RolePlaying	230	0.0102***	-0.00191	0.00212
Games -Adventure	189	0.00428	0.000588	-0.00137
Games -Simulation	181	0.00476	0.00699**	-0.00121
Games -Racing	126	-0.0146*	-0.00167	0.00396

Role of Competition

So far, we observe that privacy label disclosure negatively effects the demand (positively effecting the ranking). One could thus conjecture that disclosure by competition would positively affect the demand. To investigate the role of competition disclosure on demand, we consider two additional variables of interest – number of apps within category that are disclosing privacy labels, and the average number of datatypes being disclosed by the apps competing within category. Thus, we estimate the coefficients for the following empirical model:

$$y_{it} = \beta_0 + \beta_1 X_{it}^{dtt} + \beta_2 X_{it}^{dnot} + \beta_3 X_{it}^{dlin} + \beta_4 \sum\nolimits_{i\neq j} X_{jt}^{dtt} + \beta_5 \sum\nolimits_{i\neq j} X_{jt}^{dnot} + \beta_6 \sum\nolimits_{i\neq j} X_{jt}^{dlin} + \beta_7 X_{it} + \delta_i + \delta_{i\in C} + \varepsilon_{it} \qquad (4)$$

In the above empirical models, y_{it} is the rank of an app (i) at time (t), δ_i is the app (i) fixed effect, δ_c is the app category (c) fixed effect, and ε_{it} is the error term. X_{it}^{dttu}, X_{it}^{dnot}, and X_{it}^{dlin} are dummy variables representing the privacy labels disclosed by app (i) after the update, and X_{jt}^{dttu}, X_{jt}^{dnot}, and X_{jt}^{dlin} are continuous variable counting the number of competing apps (j) that have disclosed each of the privacy label during the same time (t). For this analysis we only consider Boolean values of disclosure by focal apps (i) and sum of Boolean values of disclosure by competing apps (j). The model estimates are presented in table 6 below.

Table 6: Regression Models for Disclosure by Competition

Ln(rank_cat)	TOP FREE APPS	TOP GROSSING APPS	TOP PAID APPS
TrackYou (bool)	-0.00860* (0.00465)	-0.00176 (0.00352)	0.00365 (0.00497)
LinkYou (bool)	0.00393 (0.00366)	0.00163 (0.00229)	-0.00140 (0.00488)
NotYou (bool)	0.00736*** (0.00282)	-0.00206 (0.00141)	-0.00439 (0.00274)
TrackYou (comp count)	0.000187* (0.000101)	0.000141** (6.86e-05)	0.000691 (0.000640)
LinkYou (comp count)	0.000546*** (9.53e-05)	0.000139* (7.21e-05)	0.000497 (0.000751)
NotYou (comp count)	-0.000158 (9.73e-05)	-4.94e-05 (6.89e-05)	0.000489** (0.000197)
_cons	3.045*** (0.198)	3.622*** (0.188)	3.204*** (0.215)
Category FE	yes	yes	yes
App FE	yes	yes	yes

Counter-intuitively, the role of competition is negative on demand - positive coefficients for TrackYou (comp count) and LinkYou (comp count). This suggests that when more competing apps are disclosing the privacy labels, non-disclosure by a focal app hurts its demand by not gaining trust of customers. As a next step we intend to investigate possible non-linear nature of this relationship where disclosure by small number of competing apps may have a no (or positive) effect on demand of focal app but as more apps start disclosing the privacy labels to create a norm in category, non-disclosure by focal app negatively impacts its demand.

Conclusion

The question of whether privacy labels will impact market for these apps is an intriguing question. Using a rich dataset comprising of 37,088 apps with 2,467,154 observations over the period of 1 year, we find that privacy label disclosure impacts the demand negatively. Further analysis of data allowed us to find the app categories that see a large/small impact caused by privacy label disclosure. These results suggest that privacy label disclosure for some app categories could help build trust and gain additional demand whereas for some other categories, it could negatively affect the demand. Most of our findings are statistically significant for free apps, with almost no significance in case of paid apps and some mixed outcomes in case of top revenue grossing apps.

As a result, we posit that disclosure of labels that suggest collection of personal data (except in categories that need that for functionality), customers penalize those app with reduced demand. Additionally, disclosure of labels that clearly show the non-identifying and non-personal data being collected, increased demand for the Apps – possibly because of increased trust in the app. We believe that there are multiple dimensions to this issue and evaluating user demand has strong economic underpinnings. We feel that we have very robust the detailed data set and a clearly defined model. We are confident that our paper will generate interesting and relevant discussion during the conference.

References

Acquisti, A., and Grossklags, J. 2005. "Privacy and Rationality in Individual Decision Making," *IEEE Security Privacy* (3:1), pp. 26–33. (https://doi.org/10.1109/MSP.2005.22).

Adjerid, I., Peer, E., and Acquisti, A. 2018. "Beyond the Privacy Paradox: Objective Versus Relative Risk in Privacy Decision Making," *Management Information Systems Quarterly* (42:2), pp. 465–488.

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., and Turner, E. 2019. "Americans and Privacy: Concerned, Confused and Feeling Lack of Control Over Their Personal Information," *Pew Research Center: Internet, Science & Tech*, , November 15. (https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/, accessed May 3, 2022).

Barkhuus, L., and Dey, A. 2003. "Location-Based Services for Mobile Telephony: A Study of Users' Privacy Concerns," in *INTERACT*.

Bélanger, F., and Crossler, R. E. 2011. "Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems," *MIS Quarterly* (35:4), Management Information Systems Research Center, University of Minnesota, pp. 1017–1041. (https://doi.org/10.2307/41409971).

Boyles, J. L., Smith, A., and Madden, M. 2012. "Privacy and Data Management on Mobile Devices," *Pew Research Center: Internet, Science & Tech*, , September 5. (https://www.pewresearch.org/internet/2012/09/05/privacy-and-data-management-on-mobile-devices/, accessed August 23, 2021).

Garg, R., and Telang, R. 2013. "Inferring App Demand from Publicly Available Data," *Management Information Systems Quarterly* (37:4), pp. 1253–1264.

Golan, E., Kuchler, F., Mitchell, L., Greene, C., and Jessup, A. 2001. "Economics of Food Labeling," *Journal of Consumer Policy* (24:2), pp. 117–184. (https://doi.org/10.1023/A:1012272504846).

Kim, D. J., Ferrin, D. L., and Rao, H. R. 2008. "A Trust-Based Consumer Decision-Making Model in Electronic Commerce: The Role of Trust, Perceived Risk, and Their Antecedents," *Decision Support Systems* (44:2), pp. 544–564. (https://doi.org/10.1016/j.dss.2007.07.001).

Mayya, R., and Viswanathan, S. 2021. "Delaying Informed Consent: An Empirical Investigation of Mobile Apps' Upgrade Decisions," SSRN Scholarly Paper No. ID 3457018, SSRN Scholarly Paper, Rochester, NY: Social Science Research Network, March 14. (https://doi.org/10.2139/ssrn.3457018).

Molitor, D., Spann, M., Ghose, A., and Reichhart, P. 2020. "Effectiveness of Location-Based Advertising and the Impact of Interface Design," *Journal of Management Information Systems* (37:2), Routledge, pp. 431–456. (https://doi.org/10.1080/07421222.2020.1759922).

Romanosky, S., Telang, R., and Acquisti, A. 2011. "Do Data Breach Disclosure Laws Reduce Identity Theft?," *Journal of Policy Analysis and Management* (30:2), pp. 256–286. (https://doi.org/10.1002/pam.20567).

Appendix A

Privacy label categories:

- Data Used to Track You:
- Data Linked to You
- Data Linked to You
- Data Not Collected

Privacy data types in each of privacy labels (https://developer.apple.com/app-store/app-privacy-details/):

Contact Info	
Name	Such as first or last name
Email Address	Including but not limited to a hashed email address
Phone Number	Including but not limited to a hashed phone number
Physical Address	Such as home address, physical address, or mailing address
Other User Contact Info	Any other information that can be used to contact the user outside the app
Health & Fitness	
Health	Health and medical data, including but not limited to data from the Clinical Health Records API, HealthKit API, MovementDisorderAPIs, or health-related human subject research or any other user provided health or medical data
Fitness	Fitness and exercise data, including but not limited to the Motion and Fitness API
Financial Info	
Payment Info	Such as form of payment, payment card number, or bank account number. If your app uses a payment service, the payment information is entered outside your app, and you as the developer never have access to the payment information, it is not collected and does not need to be disclosed.
Credit Info	Such as credit score
Other Financial Info	Such as salary, income, assets, debts, or any other financial information
Location	
Precise Location	Information that describes the location of a user or device with the same or greater resolution as a latitude and longitude with three or more decimal places
Coarse Location	Information that describes the location of a user or device with lower resolution than a latitude and longitude with three or more decimal places, such as Approximate Location Services
Sensitive Info	

Sensitive Info	Such as racial or ethnic data, sexual orientation, pregnancy or childbirth information, disability, religious or philosophical beliefs, trade union membership, political opinion, genetic information, or biometric data
Contacts	
Contacts	Such as a list of contacts in the user's phone, address book, or social graph
User Content	
Emails or Text Messages	Including subject line, sender, recipients, and contents of the email or message
Photos or Videos	The user's photos or videos
Audio Data	The user's voice or sound recordings
Gameplay Content	Such as saved games, multiplayer matching or gameplay logic, or user-generated content in-game
Customer Support	Data generated by the user during a customer support request
Other User Content	Any other user-generated content
Browsing History	
Browsing History	Information about content the user has viewed that is not part of the app, such as websites
Search History	
Search History	Information about searches performed in the app
Identifiers	
User ID	Such as screen name, handle, account ID, assigned user ID, customer number, or other user- or account-level ID that can be used to identify a particular user or account
Device ID	Such as the device's advertising identifier, or other device-level ID
Purchases	
Purchase History	An account's or individual's purchases or purchase tendencies
Usage Data	
Product Interaction	Such as app launches, taps, clicks, scrolling information, music listening data, video views, saved place in a game, video, or song, or other information about how the user interacts with the app
Advertising Data	Such as information about the advertisements the user has seen
Other Usage Data	Any other data about user activity in the app
Diagnostics	
Crash Data	Such as crash logs
Performance Data	Such as launch time, hang rate, or energy use
Other Diagnostic Data	Any other data collected for the purposes of measuring technical diagnostics related to the app

Other Data	
Other Data Types	Any other data types not mentioned