

Usability of Privacy Control and Disclosure Mechanisms

by David G. Balash

B.S. in Computer Engineering, May 1997, Iowa State University
M.S. in Computer Science, January 2016, The George Washington University

A Dissertation Proposal submitted to

The Faculty of
The School of Engineering and Applied Science
of the George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

October 28, 2022

Dissertation directed by

Adam J. Aviv
Associate Professor of Computer Science

Usability of Privacy Control and Disclosure Mechanisms

David G. Balash

Dissertation Research Committee:

Yasemin Acar, Assistant Professor of Computer Science,
Committee Member

Adam J. Aviv, Associate Professor of Computer Science,
Dissertation Director

Lujo Bauer, Professor of Electrical and Computer Engineering,
Committee Member

Timothy Wood, Associate Professor of Computer Science,
Committee Member

Arkady Yerukhimovich, Assistant Professor of Computer Science,
Committee Member

© Copyright 2022 by David G. Balash
All rights reserved

Abstract

Usability of Privacy Control and Disclosure Mechanisms

Rapid and unrestrained technological advancement has brought with it many new challenges for the privacy of individuals. Most people want control over the access and use of their personal information, but many feel resigned to the loss of privacy in the face of ubiquitous data collection. Privacy matters in an increasingly digital world; it protects individual autonomy and enables people to manage their reputations and maintain social boundaries. Furthermore, privacy limits the power of private sector companies and governments while supporting freedom of thought, speech, and political expression. User-facing privacy controls that allow people to manage and limit the data collected about them, as well as disclosure mechanisms that provide transparency about the type, amount, and reason for information collection are available, and this thesis shows users' perspectives on the usability and effectiveness of these mechanisms.

Privacy dashboards are privacy control tools that allow users of online services both to review the data collected about them and to manage their privacy settings. Privacy nutrition labels are a privacy disclosure mechanism that aims to simplify and standardize communication of privacy behavior, similar to food nutrition labels. This thesis will seek to understand these methods of disclosure and control; in particular we focus on (i) users' security and privacy perceptions of tools, such as privacy dashboards and permissions control pages, (ii) users' understanding of advertisement inferences made about them based on their past online activity, (iii) the app-based privacy nutrition label ecosystem, and (iv) the impact of privacy labels on users' risk perception and willingness to install apps.

We show that while privacy dashboards, particularly Google's My Activity dashboard, may increase awareness of data collection, the net result may be that dashboard interactions decrease perceived concern for data collection while at the same time increasing perceptions of the benefit of data collection. This is likely due both to the overwhelming nature of

managing so much collected data via the dashboard, and to the likelihood that reviewing past interactions with Google products may remind users of the benefits and trust they have in Google to provide such services. Motivated by the fact that privacy dashboards tend only to display the raw data collected and not necessarily how that data is used, we performed an interactive user study ($n = 174$) where participants both reviewed a selection of their own activities they performed on Google and assigned advertising interests they believe that Google learns about them from those activities. The goal is to determine if helping users “connect the dots” from data collection to how that data could be used to infer their advertising interests impacts perceived benefit of and concern for online data collection, as well as to assess how accurate users are at this task and how they react to what advertising interests that Google actually applies to them.

To better understand the privacy label ecosystem, the categories and types of data collected, and the purposes that developers used to justify that collection, we collected and analyzed 36 weekly snapshots of 1.6 million apps between July 15, 2021 and March 17, 2022. Of apps with labels, 17.3% collect data used to track users, 37.6% collect data that is linked to a user identity, and 41.9% collect data that is not linked. Only 42.1% of apps with labels indicate that they do not collect any data. We found that because many apps indicate that they do not collect any data—even apps that would seem likely to collect or link data—trusting the veracity of privacy labels is still an open question. However, privacy labels have the potential to help users make informed choices when selecting an application to install. We will also research user perceptions, understanding, and behavior with regards to the app-based privacy labels. This is an opportunity to understand user behavior regarding privacy communication provided in a new format and to evaluate the impact of that communication on users’ decisions about whether or not to install an application.

Table of Contents

Abstract	iv
List of Figures	viii
List of Tables	x
1 Introduction	1
1.1 Background: Privacy Control Mechanisms	1
1.2 Research Task: Understanding the Usability of Privacy Dashboards	2
1.3 Research Task: Understanding the Usability of Privacy Controls	3
1.4 Background: Privacy Disclosure Mechanisms	4
1.5 Research Task: Understanding App-based Privacy Nutrition Label Ecosystems	5
1.6 Research Task: Understanding App-based Privacy Nutrition Label Perceptions	6
1.7 Research Questions	6
1.8 Organization	8
2 Related Work	11
2.1 Privacy Dashboards	11
2.2 Ad Settings and Inferences	13
2.3 Application Programming Interface Privacy	16
2.4 App-Based Privacy Nutrition Labels	18
3 Evaluating User Perceptions and Reactions to Google’s My Activity	22
3.1 Background: Google My Activity	24
3.2 Method	25
3.2.1 Study Procedure	25
3.2.2 Recruitment and Demographics	28
3.2.3 Analysis Methods and Metrics	28
3.2.4 Ethical Considerations	30
3.2.5 Limitations	30
3.3 Results	31
3.3.1 RQ1: Awareness and Understanding	31
3.3.2 RQ2: Impact on Benefit and Concern	38
3.3.3 RQ3: Behavioral Change	45
3.4 Discussion	51
3.5 Conclusion	54
4 Security and Privacy Perceptions of Third-Party Application Access for Google Accounts	56
4.1 Method	58
4.2 Results	63
4.2.1 Measurements	63
4.2.2 Awareness and Understanding	65

4.2.3	Granting and Reviewing Account Access	74
4.2.4	Reflection and Features	80
4.3	Discussion and Conclusion	85
5	Longitudinal Analysis of Privacy Labels in the Apple App Store	89
5.1	Background	91
5.2	Methodology	94
5.3	Overall App Store Trends	96
5.4	Comparing App Metadata and Privacy Labels	100
5.5	Privacy Label Adoption and Changes	107
5.6	Discussion and Conclusion	112
6	Proposed Work and Timeline	116
6.1	The Impact of App-Based Privacy Labels on Users	116
6.2	Ad Settings Study	123
	Bibliography	129
A	Google My Activity	144
A.1	Screening Survey Instrument	144
A.1.1	Main Survey Instrument	144
A.2	Qualitative Codes	151
A.3	Demographics	158
A.4	Screenshots of My Activity and the Survey	159
B	API Privacy	163
B.1	Survey Instrument	163
B.1.1	First Survey	163
B.1.2	Second Survey	166
B.2	Additional Figures and Tables	174
B.3	Qualitative Codebook	183
C	Longitudinal Analysis of Privacy Labels	188
C.1	Additional Figures and Tables	188
D	Advertisement Inferences	209
D.1	Survey Instrument	209
E	User Perceptions of App-Based Privacy Labels	216
E.1	Survey Instrument	216

List of Figures

3.1	Main study protocol	25
3.2	Activity awareness bar plot	32
3.3	Remember activity bar plot	33
3.4	Better understanding after visiting My Activity	35
3.5	Collecting activity is appropriate bar plot	36
3.6	Level of concern before and after visiting My Activity	37
3.7	Alluvium plots of level of concern and frequency of benefit	39
3.8	Change Settings, Review Activities, and Change Behavior Bar Plot	47
3.9	Paid plan bar plot	51
4.1	Before granting access bar plot	64
4.2	Aware recall keep bar plot	66
4.3	Last time using app bar plot	68
4.4	Access benefits, concerns, and want to change bar plot	69
4.5	How confident permission bar plot	70
4.6	How necessary permission bar plot	71
4.7	How concerned permission category bar plot	72
4.8	Before granting access bar plot	74
4.9	How often do you review services bar plot	78
4.10	Change settings and review apps bar plot	81
4.11	Reminder to review, require reapproval bar plot	83
4.12	Approve every block data bar plot	84
5.1	Privacy Label With Three Privacy Types	92
5.2	Privacy Label Details	92

5.3	Anatomy of a Privacy Label	93
5.4	Num. of apps vs. Num. Privacy Types	96
5.5	Privacy Type Venn Diagram	97
5.6	Purposes By Privacy Type Ratio	100
5.7	Data Category By Privacy Type Ratio	101
5.8	Data Type By Privacy Type Ratio	102
5.9	Content Rating By Privacy Label Type Ratios	103
5.10	Rating Counts By Privacy Label Type Ratios	103
5.11	App Size By Privacy Label Type Ratios	106
5.12	App Costs By Privacy Label Type Ratios	106
5.13	First Compliance	108
5.14	Privacy Type Shifts	109
A.1	My Activity user interface	159
A.2	My Activity item details	160
A.3	Exploring My Activity during the survey	161
A.4	Activity presentation in the survey	162
B.1	Google account access authorizations.	180
B.2	Google’s “Apps with access to your account” page.	181
B.3	Dropbox for Gmail third-party app as displayed on Google’s “Apps with access to your account” page.	182
C.1	Release Date By Privacy Label Type Magnitude	188
C.2	App Genre By Privacy Label Type Ratios	189
C.3	Category by Purpose Data Linked to You Heatmap	189
C.4	Category by Purpose Data Not Linked to You Heatmap	190
C.5	Privacy Type Change Across Runs	190

List of Tables

1.1	Studies that make up this thesis.	10
3.1	Participant demographics	29
3.2	Post-exposure concern regression analysis	42
3.3	Post-exposure frequency of benefit regression analysis	43
3.4	Review activities regression analysis	48
3.5	Use Google differently regression analysis	49
4.1	Characteristics of the study participants	61
4.2	Remove app regression analysis	79
5.1	Changes observed in the addition and removal of data categories from privacy labels. Twice as many apps that changed the composition of their labels added a data category as those that removed a data category.	110
6.1	Detailed Analysis Plan.	120
6.2	Detailed Analysis Plan Continued.	121
6.3	Timeline for proposed tasks.	122
6.4	Detailed Analysis Plan.	126
6.5	Detailed Analysis Plan Continued.	127
6.6	Timeline for proposed tasks.	128
A.1	Full demographics	158
B.1	Full demographics data of the participants of the first survey.	174
B.2	Full demographics data of the participants of the second survey.	175

B.3	The top 26 (4 apps tied for 23rd ranked by authorized count) apps with access, the categories for the requested permissions, the average number of days authorized to the participants' Google account, and the number of permissions each app requested.	176
B.4	The top 25 authorized SSO ranked by the number of authorized accesses to participants' Google accounts, and the number of permissions each SSO requested.	177
B.5	The top 28 (5-way tie for permissions count of 8) most requested permissions by third-party apps with authorized access to participants' Google accounts, the permission category, and a count of how many of each permissions were authorized.	178
B.6	The top 25 most requested permissions by SSO with authorized access to participants' Google accounts ranked by the number of each permission authorized, and the permission category.	179
C.1	Privacy Type Shifts in correlation with version updates.	191
C.2	Observed shifts in privacy types associated with each data category.	192
C.3	Observed shifts in purposes associated with each data category.	196

Chapter 1: Introduction

This thesis seeks to better understand the usability of privacy control and disclosure mechanisms from the perspective of users. Specifically we (i) investigate users' security and privacy perceptions of tools such as privacy dashboards and permissions control pages, (ii) examine users' understanding of advertisement inferences made about them based on their past online activity, (iii) measure and analyze the app-based privacy nutrition label ecosystem, and (iv) determine the impact of privacy labels on users' risk perception and willingness to install apps.

1.1 Background: Privacy Control Mechanisms

The European Union's (EU) privacy law, the General Data Protection Regulation (GDPR) [110], was put into effect in 2018. Organizations outside of the EU must also comply with the law if they target or collect data related to people in the EU. The GDPR requires organizations to provide "access to the personal data" (Article 15), that notices be provided in "a concise, transparent, intelligible and easily accessible form, using clear and plain language" (Article 12), that organizations provide privacy choices including "erase personal data without undue delay" (Article 17), and the ability to opt-out where "personal data are processed for direct marketing purposes" (Article 21).

Privacy regulation in the United States is provided by sector-specific federal laws and state privacy laws. For example, the Children's Online Privacy Protection Act (COPPA) [25] enacted by the U.S. Congress in 1998 requires companies to comply with parental requests regarding data collection and deletion for children under 13, and the California Consumer Privacy Act (CCPA) [107] passed in 2020 provides California residents the right to opt out of the sale of their personal information by companies. The CCPA was subsequently amended to include rights such as "Consumers should know who is collecting their personal

information, how it is being used, and to whom it is disclosed.” and “Consumers should have meaningful control of their personal information, including their sensitive personal information, and meaningful options over how it is collected, used, and disclosed.”

Data-privacy laws and regulations, like GDPR and CCPA, have led to companies providing increased data collection management tools due to data access requirements. We investigate the usability of these data collection management tools.

In order to comply with the regulatory “right of access” businesses began to offer so-called data downloads [146], which are files or archives of files containing data collected about a user. Despite the access data downloads provide to user data maintained by companies, usability limitations keep them from providing users with meaningful transparency. Data downloads may be large archives containing gigabytes of data, many of them containing files in formats such as JSON and CSV, using UNIX timestamps (seconds since epoch, e.g. 1661452915).

Subsequent to data downloads, businesses began to provide web based transparency tools such as privacy dashboards. Privacy dashboards are interactive web pages that allow users of online services to review and manage data collection and inform their future sharing decisions. While there has been research suggesting privacy dashboards [51, 73, 119, 159] increase users’ understanding of data collection, particularly around online behavioral advertising [22, 115, 143, 151, 152] and interest inferences [38, 116, 141], this thesis investigates the impact of privacy dashboards on the perceived risks or benefits of the data collection itself.

1.2 Research Task: Understanding the Usability of Privacy Dashboards

Since 2016, Google has offered a privacy dashboard, My Activity, which allows users to review and delete their activity data from Google services. Google defines My Activity as “a central place to view and manage activity such as searches you’ve done, websites you’ve visited, and videos you’ve watched.” Depending on how and what Google products they

use, My Activity provides a more or less comprehensive picture of a user’s digital routine. Ranging from what apps users started on their Android devices, over what websites they’ve visited while using Chrome, to what places they’ve visited or looked up using Google Maps. Each activity consists of a set of metadata like a timestamp, type of Google product, and a descriptive label often containing an URL pointing to the Google service or website in question.

We conducted an online survey with $n = 153$ participants to understand if Google’s My Activity, as an example of a privacy transparency tool, increases or decreases end-users’ concerns and benefits regarding data collection. While most participants were aware of Google’s data collection, the volume and detail was surprising, but after exposure to My Activity, participants were significantly more likely to be both less concerned about data collection and to view data collection more beneficially. Only 25% indicated that they would change any settings in the My Activity service or change any behaviors. This suggests that privacy dashboards are quite beneficial for online services as they garner trust with their users and improve their perceptions without necessarily changing users’ behaviors.

1.3 Research Task: Understanding the Usability of Privacy Controls

Online services like Google provide a variety of application programming interfaces (APIs). These online APIs enable authenticated third-party services and applications (apps) to access a user’s account data for tasks such as single sign-on (SSO), calendar integration, and sending email on behalf of the user, among others. Despite their prevalence, API access could pose significant privacy and security risks, where a third-party could have unexpected privileges to a user’s account.

To gauge users’ perceptions and concerns regarding third-party apps that integrate with online APIs, we performed a multi-part online survey of Google users. First, we asked $n = 432$ participants to recall if and when they allowed third-party access to their Google account: 89% recalled using at least one SSO and 52% remembered at least one third-party

app. In the second survey, we re-recruited $n = 214$ participants to ask about specific apps and SSOs they’ve authorized on their own Google accounts. We collected in-the-wild data about users’ actual SSOs and authorized apps: 86% used Google SSO on at least one service, and 67% had at least one third-party app authorized. After examining their apps and SSOs, participants expressed the most concern about access to personal information like email addresses and other publicly shared info. However, participants were less concerned with broader—and perhaps more invasive—access to calendars, emails, or cloud storage (as needed by third-party apps). This discrepancy may be due in part to trust transference to apps that integrate with Google, forming an implied partnership. Our results suggest opportunities for design improvements to the current third-party management tools offered by Google; for example, tracking recent access, automatically revoking access due to app disuse, and providing permission controls.

1.4 Background: Privacy Disclosure Mechanisms

Privacy policies and terms of service agreements, free-text explanations of what data services collect and how that data is used, have become a standard and accepted part of notice and consent laws, and failure to provide an accurate and comprehensive privacy policy could lead to serious legal consequences. Companies are well incentivized to provide broad privacy policies that provide legal cover for their data collection practices in a way that protects them from any jeopardy, including hiring lawyers and other policy experts to craft and review them. Given their length and legal jargon, research continually shows that privacy policies are neither well understood [122] nor actively reviewed by most users [79].

One proposed alternative to privacy policies that is intended to improve privacy communication is to do away with natural language presentations of privacy behavior and instead use *privacy nutrition labels* [82] or, more simply, *privacy labels*. These are prescriptive labeling of applications modeled after food nutrition labels [54]. Like a food label, a privacy label describes the data collection and usage practices of a service and have been proposed

for a range of products, most notably for the Internet of Things (IoT) [41, 83, 85], where interfaces and interactions can be limited. The key idea is that labels provide more clarity and transparency that is difficult to achieve via the privacy policy.

In December of 2020, Apple began requiring privacy labels [13] for all new and updated apps in the App Store. Apple’s privacy labels ask developers to self-label (without verification) the data collection and sharing practices of their apps, the purposes, the types of data, and if that data is linked to user identities. Essentially, privacy labels standardizes the presentation of privacy behavior that was previously described in the natural language text of the privacy policy.

1.5 Research Task: Understanding App-based Privacy Nutrition Label Ecosystems

We collected and analyzed 36 weekly snapshots of 1.6 million apps between July 15, 2021 and March 17, 2022 to better understand the privacy label ecosystem, the categories and types of data collected, and the purposes that developers used to justify that collection. Nearly two years after privacy labels launched, only 60.5% of apps have privacy labels, increasing by an average of 0.5% per week, mostly driven by new apps rather than older apps coming into compliance. Of apps with labels, 17.3% collect data used to track users, 37.6% collect data that is linked to a user identity, and 41.9% collect data that is not linked. Only 42.1% of apps with labels indicate that they do not collect any data. As many apps still do not have labels and those apps that have assigned labels do not typically change their labels, it appears that privacy labels are a “set and forget” mechanism that may not actually provide users with the clarity needed to make informed privacy decisions. We found that because many apps indicate that they do not collect any data, even apps that would seem likely to collect or link data, trusting the veracity of privacy labels is still an open question.

1.6 Research Task: Understanding App-based Privacy Nutrition Label Perceptions

We seek to understand how app-based privacy nutrition labels influence users' perceptions of an application. Using quantitative data collection and analysis techniques we will determine the privacy attributes and corresponding values that most impact a users' risk perception and willingness to install an application on their mobile device. Furthermore, using qualitative data collection we will attempt to gain insights into why users were influenced or not influenced by the privacy label attributes when considering privacy risks and willingness to install. Subsequently, we will use this new understanding to create recommendations for effectively communicating privacy and security risks to users on an app-based privacy nutrition label.

1.7 Research Questions

This thesis proposal addresses the following research questions:

RQ1 What are users' security and privacy perceptions of privacy dashboards and permissions control pages?

We show that while privacy dashboards, particularly Google's My Activity dashboard, may increase awareness of data collection, the net result may be that dashboard interactions *decreases* perceived concern for data collection while at the same time *increasing* perceptions of the benefit of data collection. This is likely due to both the overwhelming nature of managing so much collected data via the dashboard, and that reviewing past interactions with Google product may remind users of the benefits and trust they have in Google to provide such services.

We also show that there is significant opportunity to improve how users interact with third parties with programmatic access to their accounts by helping users to identify and remove less frequently used applications and SSOs in an automated

way, or to simply revoke access after a period of disuse. For instance, Google could require regular re-approval of apps with access to users' accounts, perhaps yearly so as not to be too disruptive. Additionally, controls should be available to limit specific permissions for third-party apps. This would allow users to better limit which aspects of their Google account each app/SSO can access with respect to the benefit being provided, rather than forcing them to accept an all-or-nothing approach. We found that user data accessed through APIs raise privacy concerns as the application developer could gain access to a considerable amount of user data accumulating on the service provider account over time. This API access is not limited to the time of login or even the active use of the application that was provided access, but can take in data from the past, and can advance to monitor future activities. We also found several privacy issues including user concerns about granting access to personal information such as emails and contacts.

RQ2 What are users' security and privacy perceptions of advertisement inferences made about them based on their past online activity?

Motivated by the fact that privacy dashboards tend to only display the raw data collecting but not necessarily how that data is used, we performed an interactive user study ($n = 174$) where participants *both* reviewed a selection of their own, real activities they performed on Google and assigned advertising interests they believe that Google learns about them from those activity. The goal is to determine if helping users "connecting the dots" from data collection to how that data could be used to infer their advertising interests impacts perceived benefit of and concern for online data collection, as well as how accurate users are at this task and their reactions to what the interests that Google actually applies to them.

RQ3 What is the current state of the app-based privacy nutrition label ecosystem?

An analysis of the rate of privacy label adoption on the Apple App Store, the categories

and types of data collected, the purposes that developers disclosed as a reason for the collection, how app meta-data is associated with privacy behavior, and instances of under-reporting data collection.

RQ4 Which app-based privacy label attributes significantly influence user risk perception and willingness to install and in what ways?

Privacy labels have the potential to help users make informed choices when selecting an application to install. But it is unknown if privacy labels have the ability to convey privacy risk to users and what impact labels have on users' willingness to install an application. It is also unknown if users care about privacy risks even if they are conveyed by the privacy labels. Therefore, it is important to understand whether privacy labels lead to better privacy outcomes for users such that users' privacy expectations align with the actual behavior of the apps that they use. We will research user perceptions, understanding, and behavior, with regards to the application privacy information (app privacy labels) added to the Apple iPhone App Store in December 2020. We will ask iPhone users directly about their experiences with the privacy labels on the Apple App Store, and how these privacy labels impacted their app purchase decision making. This an opportunity to understand user behavior in regards to privacy communication provided in a new format, namely the privacy label, and its impact on whether or not to install an application on their device.

1.8 Organization

The rest of this thesis proposal is organized as follows: Chapter 2 describes related work and gives additional background on privacy disclosure and control mechanisms; user perceptions of Google's My Activity dashboard is covered in Chapter 3; Chapter 4 focuses on the security and privacy of third-party application access for Google accounts; a large scale analysis app-based privacy nutrition labels in the Apple App store is presented in Chapter 5;

Chapter 6 concludes this proposal by presenting a timeline for thesis completion.

This proposal is based in part on work. For a summary of studies that make up this thesis please refer to Table 1.1.

Table 1.1: Studies that make up this thesis.

Paper	Description	RQ#	Chap.	Publication
Are Privacy Dashboards Good for End Users? Evaluating User Perceptions and Reactions to Google’s My Activity	We conducted an online survey with $n = 153$ participants to explore how users’ concerns of and benefits from Google’s data collection are influenced by My Activity, as an exemplar privacy dashboard.	RQ1	3	30th USENIX Security Symposium (USENIX Security 21)
Security and Privacy Perceptions of Third-Party Application Access for Google Accounts	We explore how users consider security and privacy in light of third-party API access to their Google accounts given the disclosure and control mechanisms currently available.	RQ1	4	31st USENIX Security Symposium (USENIX Security 22)
Security and Privacy Perceptions of Google Advertising Inferences	We conducted an online survey with $n = 174$ participants to explore how users’ concerns of and benefits from Google’s data collection are influenced by interacting with the advertising inferences made about them from their online activities.	RQ2	6	(In progress)
Longitudinal Analysis of Privacy Labels in the Apple App Store	We collected and analyzed 36 weekly snapshots of 1.6 million apps between July 15, 2021 and March 17, 2022 to better understand the privacy label ecosystem, the categories and types of data collected, and the purposes that developers used to justify that collection.	RQ3	5	(In submission) Network and Distributed System Security Symposium (NDSS) 2023
Impact of Privacy Labels on Users’ Risk Perception and Willingness to Install iOS Apps	We will conduct a survey to test the effectiveness of each of the Apple privacy labels along two key dimensions: ability to convey risk to users and impact on their willingness to install an iOS application.	RQ4	6	(In progress)

Chapter 2: Related Work

2.1 Privacy Dashboards

Online Behavioral Advertising. Many services track online activities of their users to infer interests for targeted advertising [152]. There is much user-facing research on Online Behavioral Advertising (OBA), including targeting and personalization [72, 151], fingerprinting and tracking [19, 32, 76, 145], opting-out [68, 69, 80, 90], privacy-enhancing technologies [28, 103, 129, 156], usable privacy notices [59, 82, 127], cookie banners and consent [97, 106, 144], and also awareness, behaviors, perceptions, and privacy expectations [2, 38, 91, 95, 116, 118].

Ur et al. [143] conducted interviews to explore non-technical users' attitudes about OBA, finding that participants were surprised that browsing history can be used to tailor advertisements. Rader [115] studied users' awareness of behavioral tracking on Facebook and Google, suggesting that increased awareness of consequences of data aggregation led to increased concern. Chanchary and Chiasson [22] explored users' understanding of OBA and tracking prevention tools, noting that participants expressed more willingness to share data given control mechanism over collected data. We find similarly in this study that My Activity is such a tool: Participants expressed decreased concern with data collection and were unlikely to change collection settings.

Most recently, Wei et al. [151] studied the advertising ecosystem of Twitter, exploring ad targeting criteria. Similar to our work, participants shared some of their Twitter data via a browser extension. The authors suggested that transparency regulations should mandate that the "right of access" not only includes access to the raw data files, but also a clear description and tools to visualize the data in a meaningful way. My Activity provides such a meaningful way to visualize and access this data, but unfortunately, it still may not sufficiently motivate users to manage data collection.

Transparency and Privacy Dashboards. Transparency tools and privacy dashboards, which allow users to explore and manage data collection and privacy from online services, have been extensively proposed and explored in the literature [39, 73, 78, 103, 117, 119, 132, 141, 152, 159]. With the European General Data Protection Regulations (GDPR) (and other similar laws), data access requirements will likely lead to an increase in transparency tools and dashboards. Below we outline some of the more related work.

Rao et al. [117] suggested that dashboards were insufficient in providing transparency in to the creation of user profiles in a study of ad profiles from BlueKai, Google, and Yahoo, and as a result participants did not intend to change behaviors. This same lack of transparency in My Activity may explain why many participants do not intend to change behaviors or settings. Schnorf et al. [132] found that offering more control does not lead to less trust when exploring inferred interest transparency tools, and we find similarly with My Activity.

Angulo et al. [6] and Fischer-Hübner et al. [51] developed *Data Track*, a transparency tool for disclosing users data for different online services. Tschantz et al. [141] compared inferred values displayed in Google’s Ad Settings [65] to self-reported values, finding that logged in users were significantly more accurate. Weinshel et al. [152] developed an extension that visualizes information that trackers could infer from browsing habits, surprising users about the extent and prevalence of data collection. Our participants were aware of Google’s data collection but also surprised by its scope.

Recently, Rader et al. [116] investigated users’ reactions to Google’s and Facebook’s profile inferences, and while many participants understood inferences to be a description of past activities, they were challenged to understand them as predictive of future interests and actions. Rader et al. argued for better transparency mechanisms, by adding explanations of how inferences might get used, and restricting inferences to only include the ones that can be explained by users, and thus, are not based on aggregation or inaccurate assumptions. Meanwhile, Herder and van Maaren [73] also found that removing derived and inferred

data has a positive effect on trust and perceived risk. Note that My Activity shows raw data, not inferred data, and it may be the case that better connecting specific inferences to data collection could improve transparency and better inform user choices.

Most related to our work, Earp and Staddon [39] conducted a pilot study with about 100 undergraduate students on Google Ad Settings and Google Web History that—somewhat unfortunately—was rebuilt and became Google My Activity during their data collection in 2016. For the participants that had “sufficient” data accessible, they found no evidence that the tools were harmful to user trust and privacy. Our work confirms this finding, and goes further by showing that My Activity can be helpful in reducing concerns and increasing perceived benefits for end users. Additionally, as My Activity has been active for 4–5 years at the time of our study, our work is able to explore the impact of this transparency tool.

2.2 Ad Settings and Inferences

Inferences. Platforms such as Google and Facebook use proprietary AI systems to infer users interests for the purposes of online behavioral advertising. In response to new privacy regulations such as the E.U. General Data Protection Regulation (GDPR) and the U.S. California Consumer Privacy Act (CCPA), which require better transparency about data collection and use, these platforms have begun providing web pages where users can view their inferred ad interests. In a study of people’s reactions upon viewing their inferences on Google and Facebook, Rader et al. [116] found that users’ understanding of inferences are bounded by their own past online behaviors and their self-perceptions of their interests (i.e. things they are directly knowledgeable about), they do not envision inferences from the perspective of the platform. Most people are unaware of the kinds of inferences that AI based behavioral advertising can automatically make about them and are often unfamiliar with how inferences are derived via aggregation. To investigate inference literacy Warshaw et al. [150] interviewed 21 adults in the US and found that few believed companies can make the type of deep personal inferences that companies now routinely make with the use of AI.

Moreover, they discovered inference literacy beliefs to be clustered around two concepts: inferences based largely on directly gathered demographic data, and inferences based on straightforward processing of online behavioral data.

Opaque ad-targeting practices and behavioral profiling can create user privacy concerns, resignation, and distrust. Barbosa et al. [16] built a tool to help users better understand and engage with automatically constructed profiles that infer potential interests and demographics.

Their technology displayed in-the-moment notifications of the incremental construction of user profiles from both web browsing and activities in the physical world increasing both transparency and control of profile construction in real time. They found that this increased visibility helped users understanding of how particular actions can lead to specific ads. Participants in their study wanted to be engaged with the profile building (i.e. human in the loop) when carrying out activities that did not speak strongly to their identities, such as searching for something for a friend, or during private conversations and activities they deem sensitive, such as topics relating to health or finance. Their findings suggested that engaging users at opportunistic moments during pervasive computing with their own inferred profiles can create more trustworthy experiences. In order to understand how visualizing inference-level information about online tracking impacts users' knowledge, perceptions, and attitudes Weinshel et al [152] built a browser extension to conduct a longitudinal field study. They found that after participants had used the browser extension that visualizes examples of long-term information that third-party trackers could have inferred from their web browsing, participants had a more accurate perception of the extent of tracking and were more likely to take privacy-protecting actions.

Targeted advertising companies such as Google and Facebook generate user specific interests from users' activities while engaged with their platforms online. There is a lack of understanding of what activities, contexts, and sentiments lead to particular inferred interests. In controlled experiments created to gain insights into how Facebook generates interests

from user activities Sabir et al. [126] found that even minor activity such as scrolling through a page can lead to an interest inference, that inferred interests can often be inaccurate, and that Facebook's explanations of these inferences were often too generalized and misleading. Datta et al. [31] created an automated tool called AdFisher to explore how users behaviors impact Google's ads and the inferences found on Google's Ad Settings webpage. They found that rather than providing transparency the Ad Settings was opaque about some features of a user's profile and that visiting certain webpages can change the ads shown to users while the inferences on the settings page remained unchanged. They also found evidence of bias as Google showed fewer instances of an ad related to high paying jobs to accounts with the gender set to female. An earlier work by Wills & Tatar [155] used controlled browsing to analyze the ads shown to users and examine the inferred interests displayed in the ad preference managers provided by Facebook and Google. They also found cases in which ads were shown to users but did not appear as categories in the ad preference managers. Furthermore, they found that non-contextual ads were displayed related to inferred sensitive topics regarding sexual orientation, health and financial matters; an apparent violation of published policies of Facebook and Google.

To determine what information advertisers actually infer about users and if the inferences are accurate Bashir et al. [17] conducted a study of four ad preference managers from Google, Facebook, Oracle BlueKai, and Nielsen eXelate by gathering full interest profiles from participants with a browser extension combined with asking participants if they were actually interested in these topics. They found that participants were strongly interested in only 27% of the interests in their profiles and that browsing history only explains a small percentage of interests, suggesting that other means of tracking such as browser fingerprinting and cross-device tracking may be critical for building user interest profiles. Additionally, they found that privacy-conscious behaviors had no significant correlations with interest profile size. Andreou et al. [5] measured the Facebook advertising ecosystem and found that the median number of interests inferred for a user is 310 and that a significant number of

advertisers employ targeting strategies that could be either invasive or opaque.

2.3 Application Programming Interface Privacy

Third-party Apps and SSO Services That Use Online APIs. Russell, et al. [125] characterize online APIs as among: *content-focused APIs* that provide data; *feature APIs* that integrate existing software functionality from elsewhere; *unofficial APIs* that (unintentionally) expose internal interfaces; and *analytic APIs* that track user experiences. Here, we focus on Google’s content-focused and feature APIs that enable third-party developers to register apps with Google that can perform operations on behalf of a user. Most services, including Google’s, use the OAuth standard [26] to delegate and manage these authorizations. OAuth has been the focus of much security research [14, 139, 161], and in this paper we do not investigate the security of OAuth directly but rather user awareness and concerns for such delegations.

While we primarily focus on third-party apps, we also consider SSO services as a form of third-party apps with limited functionality. Bauer et al. [18] looked at willingness to use the SSOs of Google, Facebook and other services, finding that there were concerns with information sharing through SSO, despite messaging. We find similar concerns in our study. Ghasemisharif et al. [57] studied SSO with respect to potential for account hijacking. The authors also measured the prevalence of SSOs, finding that Facebook is the most prevalent SSO service, followed by Twitter and Google. Hu et al. [74] investigated SSOs in the context of online social networks and how apps can complete an impersonation attack. And Zhou et al. completed automated vulnerability testing of SSO on the web [158]. Here we assume that the SSO is properly implemented and instead focus on user perceptions of sharing information with third-parties via SSO services.

Prior work on third-party apps have mostly focused on the Facebook ecosystem. Felt et al. [46] examined 150 Facebook platform apps in 2008, finding that 90% of the examined apps have unnecessary access to private data. Huber et al. [75] developed a method to analyze

privacy leaks in Facebook apps at scale by leveraging client-side iframes to capture network traffic. Google third-party apps do not necessarily operate client-side. More recently, Farooqi et al. [45] used “honeypot” email addresses (i. e., auto-generated accounts on an email server that the researchers control) to detect Facebook apps inappropriately collecting and using those addresses. Such a method could also be used for Google third-party apps but was not the primary focus of this research.

Permission Management for Online APIs. Our work is also related to prior research on permission management for online APIs. Similar to Wang et al. [149], who analyzed the permissions requested by Facebook API apps at install time, we explore the permissions requested by third-party apps that integrate with Google’s API. Prior work explored a subset of these permissions on Google [124]. A lack of centralization for third-party apps means there is far from comprehensive coverage. Our work expands on this effort with in-the-wild observations of apps authorized on actual users’ Google accounts. Additionally, we collect information about participants’ levels of understanding and concerns about those permissions, as well as how long apps have been authorized on their Google accounts.

Permissions have been extensively studied in the context of smartphone apps. Notably, Felt et al. [47, 49] examined Android apps and found that one-third are overprivileged, and Wijesekera et al. [153] surveyed Android users’ perceptions of app permissions, 80% of whom wished to deny at least one permission and 35% of all app permission requests as inappropriate. With the shift towards runtime, ask-on-first-use permission requests [7, 35], Wijesekera et al. in 2017 [154] developed a classifier to predict user permission preferences by taking into account the context of the permission request. Likewise, Smullen, et al. in 2020 [137] built a classifier that demonstrated users are sensitive to the purpose of particular permissions requested by apps, and that a classifier can leverage this to assist users in deciding to grant or deny permissions. Mobile platforms have since recognized that user-granted permissions can be contextual and dynamic. Starting in Android 11 [34],

users can temporarily give an app one-time access to a sensitive API call. Additionally, Android 11 auto-revokes permissions from long-unused apps. Our results show that similar contextualization has impact on users’ perceptions of permissions, and recommend moving towards auto-review and auto-revocation models for third-party apps as a whole and for individual permissions.

Privacy and Transparency Dashboards for Online Services. Finally, this research is also related to work on privacy and transparency dashboards for online services. These have been both extensively proposed and explored in the literature [39, 44, 73, 78, 103, 117, 119, 132, 141, 152, 159]. The “Apps with access to your account” page, to which we direct participants in the survey, functions similarly to other transparency dashboards; however, this dashboard offers less functionality than other dashboards.

2.4 App-Based Privacy Nutrition Labels

Longitudinal studies of online services privacy practices. Researchers have gathered evidence of online services data collection behaviors via longitudinal measurements across various platforms [1, 42, 86, 87, 92, 109, 123, 135]. Analysis of mobile apps showed sensitive data, including PII [123], is collected and shared with third parties without user consent [87]. Data collection practices were prevalent across measurements of apps in different geographic regions [136], categories [148, 162], price brackets [70, 71, 135], and app markets [88, 99, 147].

Privacy policies. Given the prominence of data collection over the Internet, researchers and practitioners, have sought to develop better ways to facilitate privacy-relevant communication between services and end-users. Privacy Policies have been the primary method that services use to self-declare their data practices [113]. They have, however, been found to be ineffective [114], and difficult to read [30, 98]. Therefore, numerous researchers have col-

laborated with practitioners to find better ways to communicate data collection and privacy practices to the end user. Over the last two decades, W3C’s Platform for Privacy Preferences (P3P) Project [56] has made multiple attempts to standardize how online services express their privacy practices. Following these recommendations, multiple browser extensions like Disconnect [37], Ghostery [58], and Abine Blur’s DoNotTrackMe [20], developed privacy icons to help users better understand practices adopted by the websites that they were browsing. The icons were, however, found to offer limited insight to users, leaving them uncertain about what data was being collected and for what purpose [130].

Privacy permissions. Where privacy labels inform, per-missions-based models give the user control over an app’s privacy practices. Both Android and iOS require all applications to use install and/or runtime permissions [11, 35]. Prior research showed low attention and comprehension rates for install-time [50]. Felt et al. [48] further found evidence of apps over-permissioning [48]. Additionally, recent work by Reardon et al. [120] showed apps circumventing the permissions-based model by instead gaining access to data using covert and side channels. These findings suggest that requesting for user permissions *after* an app has been installed is both ineffective and insufficient.

Privacy nutrition labels. Labels have been used as an effective means to communicate information to end users on products like food (Nutrition Facts) [53] and home appliances [3, 24]. Drawing inspiration from these labels, Kelly et al. [83, 84] developed a privacy label that presents the ways in which websites collect, use, and share end-users’ personal information. This was later extended [85] in the design of a “Privacy Facts” label for mobile apps. The label detailed information that apps collect along with their intended use.

Subsequently, Emami-Naeini et al. [40, 41] developed and evaluated similar labels for Internet of Things (IoT) devices and found that users factor privacy risk perceptions into their purchase. Over the years, multiple researchers have studied and provided recommendations on designing similar privacy notices from a variety of perspectives. [15, 29, 40, 41, 83–85,

128, 138].

Prior studies of Apple’s privacy labels. Li et al. [93] interviewed 12 developers about their choice in labels, noting that developers did not completely understand the creation process for privacy labels. They reported that instances of both under-reporting and over-reporting data collection were prevalent when participating developers generated privacy labels. Kollnig et al. [89] evaluated 1,759 apps before and after they added a privacy label. They looked at instances of apps collecting an identifier for cross-device tracking, and attempted to understand the impact that privacy labels had on such collection. They found apps adopting measures to circumvent Apple’s detection of their tracking activity. Zhang et al. [157] recently investigated the usability of Apple’s Privacy Labels using semi-structured interviews with 24 iOS users. This study surfaced several potential concerns with the current implementation of privacy labels including clarity of the terse explanations provided by Apple for each label’s meaning, and the lack of awareness that the labels are even included in the App Store listings.

Similar to the work that we present, Scoccia et al. [134] analyzed a small subset ($n = 17,312$) of apps on the App Store. They captured *two* snapshots of the subset of apps, seven months apart. They observed a *decrease* in the number of apps that collect data for tracking purposes, but an *increase* in overall data collection. We find similarly in analyzing 1.6 million apps over 36 weeks.

Most relevant to our work is a similar large-scale analysis of the App Store by Li et al. [94]. They collected weekly snapshots of privacy labels on the store between April 2 and November 5, 2021. They reported that only 2.7% of apps during their collection period voluntarily added a privacy label, suggesting that inactive apps have little incentive to create privacy labels. They also found developers rarely update privacy label after initial label compliance, as did we.

While Li et al.’s collection period overlapped with ours, we collected and presented

additional details for each app, including *Data Types* and *Purposes*, which helped us evaluate each label in a complete and comprehensive manner. Our collection of app metadata included similar findings to Li et al.’s analysis of *genres*, and we further investigated correlations with *content rating*, *user rating counts*, *release dates*, *app size*, and *app price*. Our correlation reports further helped question the accuracy of these labels, report on potential COPPA violations [25], and evaluate differences between *paid* and *free* apps. We additionally verified their findings on initial compliance, while also providing fine-grained reports of updates to privacy labels, which was made possible by our collection of the entire set of data that each label provides.

Chapter 3: Evaluating User Perceptions and Reactions to Google’s My Activity

Privacy dashboards [39, 51, 73] allow users of online services to review and control data collection. Google introduced an activity dashboard called *My Activity* [66] in 2016 that allows users to view their activity history (such as searches, videos, and location data), turn off activity collection, and (automatically) delete activities from their history.

While there has been research suggesting privacy dashboards [51, 73, 119, 159] increase users’ understanding of data collection, particularly around online behavioral advertising [22, 115, 143, 151, 152] and interest inferences [38, 116, 141], there is little research on the impact of privacy dashboards on the perceived risks or benefits of the data collection itself.

In this paper, we conducted an online survey with $n = 153$ participants to explore how users’ concerns of and benefits from Google’s data collection are influenced by My Activity, as an exemplar privacy dashboard. Participants were first surveyed about their concern regarding Google’s data collection and how frequently they benefit from it, both on Likert scales and in open-ended responses. They were then directed to their Google My Activity dashboard to view their own, real, activities that Google collected about them, and then participants were again asked about their concerns/or benefits. Through these methods, we were able to ask and answer the following research questions:

RQ1 [*Awareness and Understanding*] What are users’ awareness and understanding of Google’s data collection?

Participants are generally aware of and understand why Google collects activities, citing targeted advertising, personalization, and product improvements. However, while aware of the purposes, many express surprise with the volume and detail of activities.

RQ2 [*Impact on Benefit/Concern*] How does the My Activity dashboard affect users’ concern about and perceived benefit of Google’s data collection?

Concern about Google's data collection significantly decreased, and perceived benefit significantly increased post exposure to My Activity, despite participants' qualitatively describing similar concerns and benefits before and after exposure. Ordinal logistic regression indicated that those who showed higher initial concern were much more likely to reduce their concern, and across all initial benefit levels, participants were almost always likely to increase their perceived benefit.

RQ3 [*Behavioral Change*] What settings and behaviors would users change due to exposure to My Activity?

Most participants describe that they would not (37 %) or are unsure if (26 %) they would change any My Activity settings, and only 25 % indicated that they plan to use Google products differently. Logistic regression suggests that those that had an increase in concern and decrease in benefit were much more likely ($11.3\times$ and $2.1\times$, respectively) to use Google differently.

These results suggest that privacy dashboards and transparency tools are a net positive for online services. Google's My Activity both decreases concerns about and increases perceived benefit of data collection, but it is not clear that these dashboards help end-users, broadly, to increase their privacy. Most participants indicated that they would not use the features of the dashboard nor change their behavior.

This may be because many users are already privacy resigned, believing that data collection will occur regardless of their choices, or it may be that the burden of properly managing their privacy is too high despite the availability of the transparency tool. As more and more transparency tools become available, this burden will only increase, and so research into mechanisms to consolidate and automate management of data collection may greatly benefit users.

3.1 Background: Google My Activity

Google introduced *My Activity*¹ in June 2016 [108], and it enables users to manage their Google Web & App, Location, and YouTube history and other data collected from Chrome, Android, etc. My Activity is designed as a transparency tool, privacy dashboard, and data collection control mechanism and is the successor of Google’s *Web History*.

The My Activity pages offers a number of user benefits to data collection. For example, “*more personalized experiences across all Google services,*” and it offers users “*faster searches, better recommendations,*” “*personalized maps, recommendations based on places you’ve visited,*” and “*better recommendations, remember where you left off, and more.*”²

My Activity lists *activities* such as, “Searched for *USENIX 2021*,” and activity details, such as type of activity, timestamp, and device. Viewed as a single event, bundle of events, or filtered by date ranges and services, users can review or delete activities, as well as enabled/disabled data collection and ad personalization. Users see a popup when disabling activity collection warning that this action will also disable personalization and not delete previously collected data. (See *Explore My Activity* section in Appendix A.1.1 for a visual.)

In May 2019, Google added a setting to enable automatic deletion of activities (after 3 or 18 months) [102], and in August 2019, Google introduced an option to disable collecting audio recordings [21]. In June 2020, Google updated their policy to give the option for auto-deleting activities during account creation for *newly created* accounts after 18 months for Web & App and Location activities and after 36 months for YouTube activities. However, *existing* accounts will still need to proactively enable the feature [104].

¹Google’s My Activity, available at: <https://myactivity.google.com>, as of October 19, 2022.

²My Activity activity controls, available at: <https://myactivity.google.com/activitycontrols>, as of October 19, 2022.

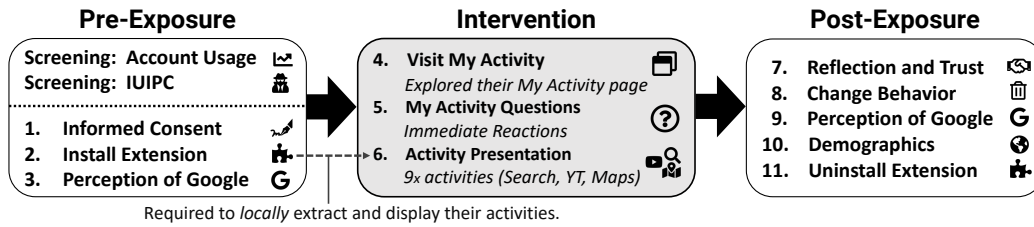


Figure 3.1: Main Study: The study was divided into three parts. During the intervention part, participants visited their own My Activity page and were questioned about nine of their activities (three per category) from Google Search, YouTube, and Maps.

3.2 Method

We designed our study for participants to directly interact with their own activity history on My Activity, following a pre-post-study design. First, participants answered questions regarding their concern for and benefit from Google’s data collection, and after exposure to My Activity, they answered the same set of questions. In the rest of this section, we outline our study protocol, recruitment, limitations, and ethical considerations.

3.2.1 Study Procedure

To ensure that participants had active Google accounts, we used a two-part structure with a *screening survey* where qualified participants were asked to participate in the *main study*. The full screening survey can be found in Appendix A.1, and the main study can be found in Appendix A.1.1.

Screening Survey We used the following inclusion criteria to screen participants for the main study: (i) the participant has an active Google account, (ii) the participant has used their Google account for more than three years, (iii) the participant currently uses Google Search, Google Maps, and YouTube.

In the screening survey we also asked participants if they have a Gmail account (as surrogate for a Google account), the age of the account, and what other Google products (besides Gmail) they use and their frequency of use and overall importance. Participants

also answered the Internet users' information privacy concerns (IUIPC) questionnaire, as described by Malhotra et al. [96], to gain insights into participants' privacy concerns.

Main Study If participants qualified they were invited to complete the main study which is divided into three stages: (i) a pre-exposure stage, in which participants install the survey browser extension that aided in administering the survey and answer questions about their perceptions of Google; (ii) an intervention stage consisting of two steps; (a) an exploration phase step and (b) an activity presentation step (iii) a post-exposure stage. To facilitate the study, we designed a custom browser extension that locally analyzes My Activity to collect aggregated information about the number of activities of users and also to fill-in survey questions. Participants are given detailed instructions to both install and uninstall the extension.

Below, we describe each part of the study in detail (see Figure 3.1 for a visual).

1. Informed Consent: Participant consented to the study; the consent included that participants would be asked to install a web browser extension and answer questions about their experience with Google's My Activity page.
2. Install Extension: Participants installed the browser extension that assisted in administering the survey. The extension also recorded aggregate information about the survey participants' number of activities per month for each activity category (e. g., Google Search, YouTube) and the date of the oldest activity, as a proxy for account age.
3. Pre-Exposure Perceptions of Google: Participants were asked about their awareness of Google's data collection practices, their level of concern, and how often they benefit from Google's collection of their online activities, both on a Likert scale and in open-ended responses. We also asked participants if they employed any strategies to limit the amount of data that Google may collect about them. The questions about perceived level of concern and benefit serve as a pre-exposure baseline and are asked again after exposure to the Google My Activity page and recent/historical Google activities. Questions: **Q1–Q4**.

4. Visit My Activity: We provided participants with a brief descriptive introduction to the My Activity service and the term “activities” as used by Google. Participants were presented with a “Sign in with Google” button and were instructed to login to their primary Google account. Then participants explored their My Activity for two minutes, managed by the browser extension with an overlay banner and restricting navigation away from My Activity. After two minutes, participants were directed back to the survey.
5. My Activity Questions: Participants were asked to provide their immediate reactions to My Activity and their reasoning for why Google is collecting this data. Participants were also asked if they perceive the data collection to be beneficial or harmful, if they have any concerns, and whether this data collection improves their experience using Google services. Questions: **Q5–Q9**.
6. Activity Presentation: Next the browser extension locally displayed three recent activities (randomly selected from 2 to 12 days old), three three-month-old activities (randomly selected from 90 to 100 days old), and three 18-month-old activities (randomly selected from 540 to 550 days old). The participants reported their awareness and recall of each of the nine activities, which were selected with an even distribution from the services Google Search, YouTube, and Google Maps. Questions: **Q10–Q14**.
7. Reflection and Trust: We then asked the participants to reflect on their post-exposure feelings and on the appropriateness of the data collection. Questions: **Q15–Q19**.
8. Change Behavior: Participants were asked what behavioral change they would likely implement after learning about My Activity, if they planned to change how long Google stores their activities, or if they would like to delete their activities. Participants were also asked if they plan to change their My Activity settings and if they would interact differently with Google products in the future. Questions: **Q20–Q25**.
9. Post-Exposure Perception of Google: We again asked participants about their concern for and benefit from Google’s data collection. Questions: **Q26, Q27**.
10. Demographics: Participants were asked to provide demographic information, such as

age, identified gender, education, and technical background. Questions: **D1–D4**.

11. Uninstall Extension: Upon completing the survey participants were instructed to remove the browser extension.

3.2.2 Recruitment and Demographics

We recruited 669 participants via *Prolific*³ for the screening survey. After applying our inclusion criteria, 447 participants qualified for the main study. Of those, 153 completed the main study; unfortunately, rates of return to the main study fell below 50 %. On average, it took 4 minutes for the screening survey and 26 minutes for the main study. Participants who completed the screening survey received \$0.50 USD and \$3.75 USD for completing the main study.

We sought a balanced recruitment between gender and five age ranges (18–24, 25–34, 35–44, 45–54, 55+) with a median participant age of 38. Purposive sampling was performed using Prolific’s built in study inclusion criteria which allows researchers to specify availability based on Prolific’s pre-screened demographics. The identified gender distribution for the main study was 52 % men, 46 % women, and 2 % non-binary or did not disclose gender. Participant demographics are presented in Table 3.1 (additional demographic information can be found in Appendix A.3).

3.2.3 Analysis Methods and Metrics

Qualitative Coding We conducted qualitative open coding to analyze 19 free-response questions. A primary coder from the research team crafted a codebook and identified descriptive themes by coding each question. A secondary coder coded a 20 % sub-sample from each of the free-response questions over several rounds, providing feedback on the codebook and iterating with the primary coder until inter-coder agreement was reached (Cohen’s $\kappa > 0.7$). We report the number of responses receiving a code and percentage of

³Prolific participant recruitment service: <https://www.prolific.co>, as of October 19, 2022.

Table 3.1: Demographic data of the participants. Age and gender data for our screening survey was provided by Prolific. The IUIPC data was collected at the end of the screening survey. Note: Prolific only provides binary gender data. To get more precise data, we asked for gender and age at the end of the main study.

		Screening (<i>n</i> = 669)		Main Study (<i>n</i> = 153)	
		n	%	n	%
Gender	Woman	317	47	71	46
	Man	344	51	79	52
	Non-binary	–	–	2	1
	No answer	8	1	1	1
Age	18–24	126	19	29	19
	25–34	152	23	35	23
	35–44	144	22	31	20
	45–54	128	19	29	19
	55+	116	17	28	18
	No answer	3	<1	1	1
		Avg.	SD	Avg.	SD
IUIPC	Control	5.8	1.0	5.9	1.0
	Awareness	6.3	0.8	6.4	0.8
	Collection	5.3	1.2	5.6	1.1
	IUIPC Combined	5.9	0.8	5.7	0.9

responses assigned that code. Note that responses may be assigned multiple codes.

Statistical Tests and Regression Analysis We performed two Wilcoxon signed-rank tests for repeated measurements on the Likert responses to the pre and post-exposure questions on concern (**Q2**, **Q26**) and benefit (**Q3**, **Q27**). The same tests were used to find differences between the responses **Q11–Q14** for the presented activities, and then post-hoc, pairwise analysis using again Wilcoxon signed-rank tests between categories, with Holm correction for overlapping measures.

We also performed two proportional odds logistic regressions to analyze which factors, in addition to the intervention, that may have influenced the Likert responses moving up or down the scales for concern (**Q26**) and benefit (**Q27**).

Finally, we performed three binomial logistic regressions on behavior change questions: Google settings **Q23**, review/delete activities **Q24**, and use Google products differently in the future **Q25**. Since we were interested whether participants planned to take action, we

binned the *unsure* and *no* responses.

3.2.4 Ethical Considerations

The study protocol was approved by our Institutional Review Board (IRB) with approval number NCR202596, and throughout the process, we considered the sensitivity of participants' My Activity data at every step. At no point did (do) the researchers have access to participants' precise Google activities. All aspects of the survey requiring access to actual Google activity was administered *locally* on the participant's machine using the browser extension. We did not collect information about individual activities to protect participants privacy, and only report information in aggregate, e. g., the number of activities per month. All participants were informed about the nature of the study prior to participating and consented to participating in both the screening and main study. At no time did the extension nor the researchers have access the participants' Google password or to any other Google account data, and all collected data is associated with random identifiers.

3.2.5 Limitations

Our study is limited in its recruitment, particularly to Prolific users residing in the U.S. We attempted to compensate by performing purposive sampling on Prolific to balance demographic factors like age and gender, but we cannot claim full generalizability of the results. Despite this limitation, prior work [121] suggests that online studies about privacy and security behavior can approximate behaviors of populations.

Social desirability and response bias may lead to participants over describing their awareness of Google data collection as they may believe that this is the expectation of the researchers. Such biases may be most present when participants indicate if they intend to change a setting or behavior.

Our regression analysis is, unfortunately, under-powered to identify small effects as we only have 153 examples. However, the pseudo $R^2 > 0.5$ for the ordinal-logistic regression,

suggesting excellent fit; the logistic regressions have pseudo $0.25 < R^2 < 0.68$, also suggesting good fits. As a result, we have confidence that the models are describing meaningful covariants, but small effects may not be captured.

Finally, as a pre-post-study we attribute changes in concern and benefit to the intervention, namely exposure to My Activity, but we cannot rule out other factors impacting changes in concern and benefit. A randomized control trial would be needed to completely rule out other factors, but using such a methodology here is unclear because there is limited control of the display of activities and behaviors of our online participants outside of the study.

3.3 Results

This section is structured along our research questions. We first present our findings concerning the participants' awareness and understanding of Google's data collection practices. Secondly, we show the impact of Google's My Activity on the perceived concern and benefit of the participants. Finally, we discuss what actions participants plan to take as a result of interacting with My Activity.

3.3.1 RQ1: Awareness and Understanding

As part of **RQ1**, we seek to understand if participants are aware of Google's My Activity, understand the scope of Google's data collection and how that data is used.

Prior Awareness of My Activity. Even though Google introduced My Activity in 2016, only a third ($n = 55$; 36 %) of the participants indicate that they have visited their My Activity page prior to our study. We also asked the participants to assess how aware they were of Google's practice to collect data on individuals' use of their services. This first question served—together with the Questions **Q2** and **Q3** (see Appendix A.1.1)—to get a first impression of participants' attitudes towards data collection and privacy. Most participants ($n = 115$; 75 %) indicated they were at least *somewhat aware* ($n = 42$; 28 %),

Prior to seeing this activity, have you been aware that Google stored this activity?

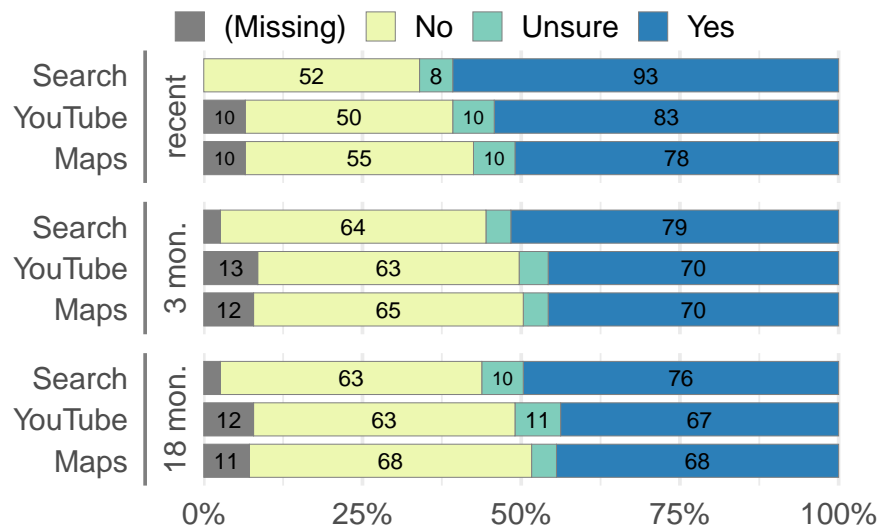


Figure 3.2: When presented with activities from their own My Activity feed, participants' awareness (Q11) seems to be similar regardless of the service. The age of the activity however has small effect on the awareness (recent against 18 months).

moderately aware ($n = 54$; 35 %), or even *extremely aware* ($n = 19$; 12 %). Only 6 (4 %) participants stated they were *not at all aware*.

Privacy Management Strategies. Qualitative coding of Q4 indicates a divide between the participants who attempt to apply a specific privacy management strategy and those who appear to be privacy resigned or unconcerned, and thus do not have a management strategy. For instance:

No strategies. I just use Chrome and whatever information Google gets they get. I signed up and accepted that they would take my data and information. (P61)

No, I don't. I don't mind that they collect data about my usage and statistics. (P21)

Half of the participants ($n = 78$; 51 %) claimed not to have strategies for managing the kind of information Google may collect about them, while 38 (25 %) participants explained that they employed web browser based strategies such as opening private or incognito

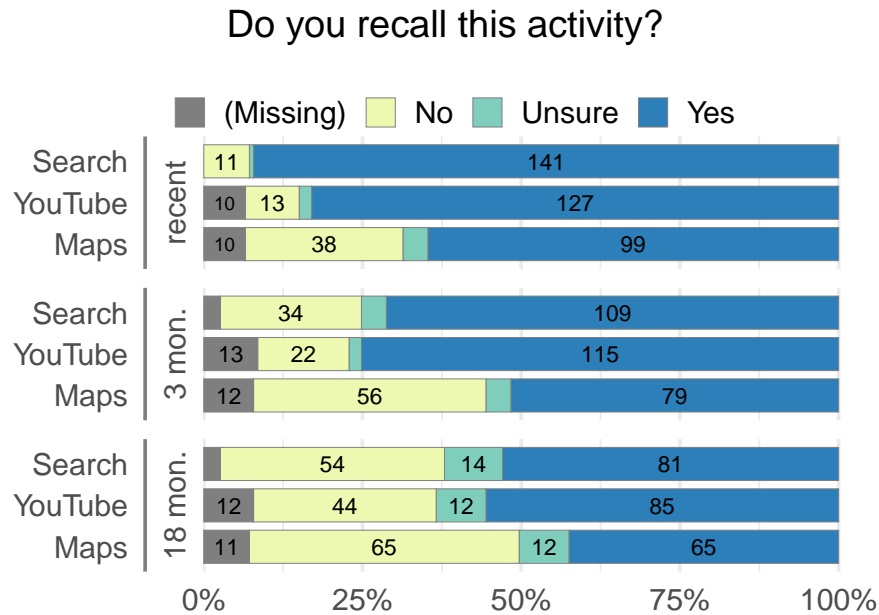


Figure 3.3: The ability of participants to recall activities (**Q10**) decreases over time independent of activity type. Google Maps activities in general seem to be harder to recall (Search / Maps: $W = 3480$; $p < 0.001$; $r = 0.25$; YouTube / Maps: $W = 3609$; $p < 0.001$; $r = 0.31$).

windows ($n = 17$; 11 %), installing ad-blocking or tracking prevention browser extensions ($n = 10$; 7 %), and clearing their browser history or cookies ($n = 9$; 6 %). Others indicated that they limit the information that they provide ($n = 25$; 16 %), limit their usage of Google products or refrain from logging into their Google accounts ($n = 7$; 5 %), provide false information ($n = 6$; 4 %), or delete information ($n = 3$; 2 %).

Scope of Data Collection. We asked a set of free-response questions after the participants visited their My Activity page to gauge immediate reactions (**Q5**). One-third ($n = 51$; 33 %) of study participants' immediate reaction was that of surprise, e. g., “I am surprised at how much of my browsing activity is saved and is identifiable” (P72), and “It’s an awful lot of my life on that page” (P11). Furthermore, 54 (35 %) participants stated that the amount of data collected on the My Activity page was more than they expected. For example:

I’m surprised at how much data google collects beside it’s own sites. I did not know it

saved the links you clicked on after a google search, for instance. (P23)

Others were not surprised ($n = 34$; 22 %) and stated the amount of data collection was as expected ($n = 30$; 20 %). For instance:

It didn't surprise me to see a tracking of all of my activity. Perhaps it gives me a way to control the information tracking in the future. (P89)

Some participants found the My Activity page helpful ($n = 16$; 11 %) and were interested ($n = 9$; 6 %), while a few participants reacted with concern ($n = 6$; 4 %), felt uncomfortable ($n = 4$; 4 %), or thought it creepy ($n = 4$; 3 %).

This is in line with closed responses to awareness of data collection types for individual activities (**Q11**); as Figure 3.2 shows, for recent search activities 61 % of the participants indicated awareness. For 18-month-old YouTube activities, only 44 % of the participants responded with *yes*. Comparing across services and activity ages, we find that there is a significant difference between awareness of recent and 18-month old activities ($W = 1511$; $p = 0.004$; $r = 0.17$).

Note that not all participants had activities for each combination of services and time frames (see missing data in Figure 3.2 and 3.3). For 24 participants, we could not obtain a full set of nine activities, 14 participants saw six activities during the survey, and six participant had seven activities. One participant saw only one activity and the remaining three participants saw two, three, or eight activities. In total 76 of 1377 records for the activity presentation were missing.

Figure 3.3 shows the results of **Q10**. The participants report higher recall for recent activities compared to older ones (recent / 3 months: $W = 1711$; $p < 0.001$; $r = 0.26$; recent / 18 months: $W = 1862$; $p < 0.001$; $r = 0.48$; 3 months / 18 months: $W = 3062$; $p < 0.001$; $r = 0.29$). Around half of the participants were able to recall their 18-month-old Search ($n = 81$; 53 %) or YouTube activities ($n = 85$; 56 %). For Maps activities the fraction was even lower ($n = 65$; 42 %). In contrast, 92 % ($n = 141$) of the participants could remember

Do you think My Activity helps you to better understand what data Google collects about you?

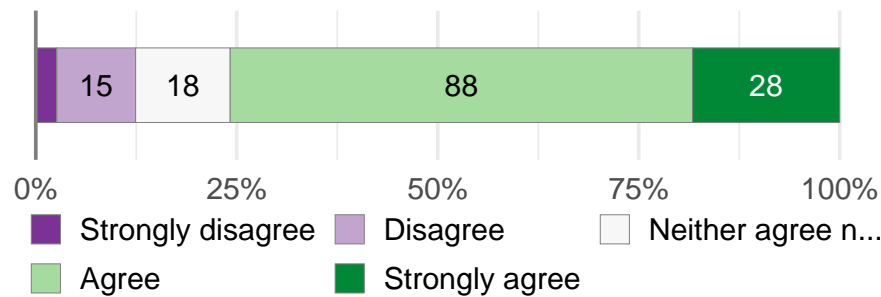


Figure 3.4: Roughly 75 % of the participants stated that My Activity helps them to better understand what data Google is collecting about them. Only around 12 % do not think My Activity aids their understanding.

their recent Google Search activities. However, even recent Google Maps activities were harder to recall for the participants ($n = 99$; 65 % could recall them). Compared with recent Google Search activities, there is a significant difference with a large effect size ($W = 2643.5$; $p < 0.001$; $r = 0.65$).

We assume this difference is due to the fact that some of the Google Maps activities were generated without the participants actively interacting with the service while Search activities are basically queries made via Google Search.

Understanding of Data Collection. We also recorded the mouse movements of the participants during their visit of the My Activity page to get an idea of whether and how they interacted with the page. We recorded an average participant scroll depth of 20553 pixels ($SD = 22285$, $min = 657$, $max = 252735$). A single activity height is approximately 200 pixels, which suggests that the average participant scrolled past approximately 100 activities during their exploration.

Asked whether My Activity helps to better understand what data Google collects, most participants ($n = 116$; 76 %) agreed. Only 12 % ($n = 19$) indicated that it did not help. Figure 3.4 shows the full results of this question. And when asked to explain why they think My Activity helps them to better understand what data Google collects (Q23_A), 61 (40 %)

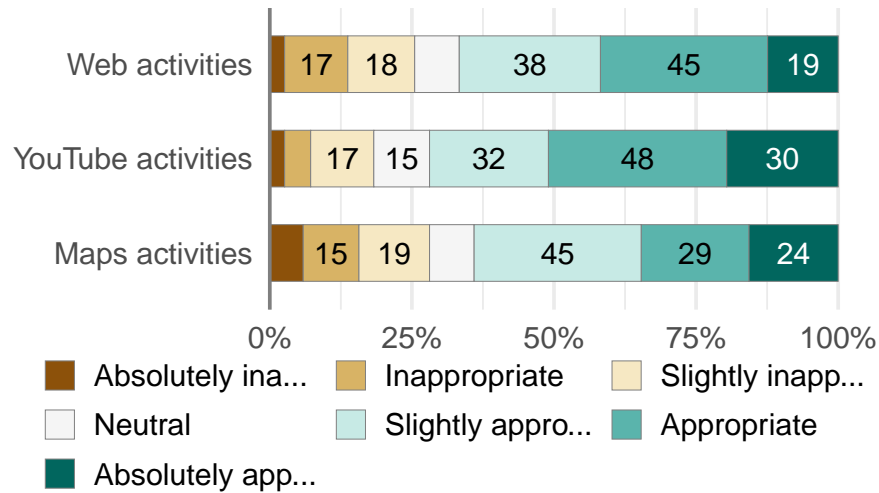


Figure 3.5: The majority of participants found the explanations Google gives as to why they collect activity data appropriate (Web: 67 %; YouTube: 72 %; Maps: 64 %).

participants reported that My Activity provides transparency about the collected data, e. g., “I didn’t realize some of this info was collected” (P4), and

I see what they are collecting. I feel like I always knew they were watching every site I visited but to quantify it gives me a better understanding. (P66)

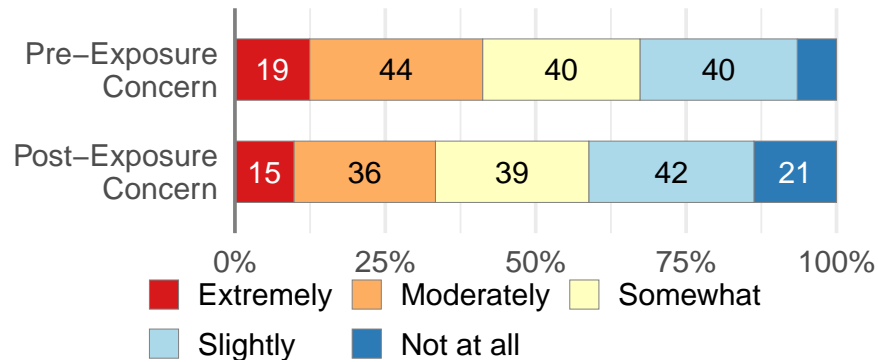
Still other participants ($n = 31$; 20 %) were skeptical and felt the My Activity page did not show all the data Google collects, e. g., “I see the data that they are retaining, but I’m concerned that there is more data being saved that they’re not sharing with me” (P148), and

I think it gives me a better understanding, but I don’t believe Google is being completely transparent on their end with what they keep or use. It is just what I can control on my end. (P69)

For some participants ($n = 13$; 8 %) My Activity did not help them better understand what data Google collects. For example:

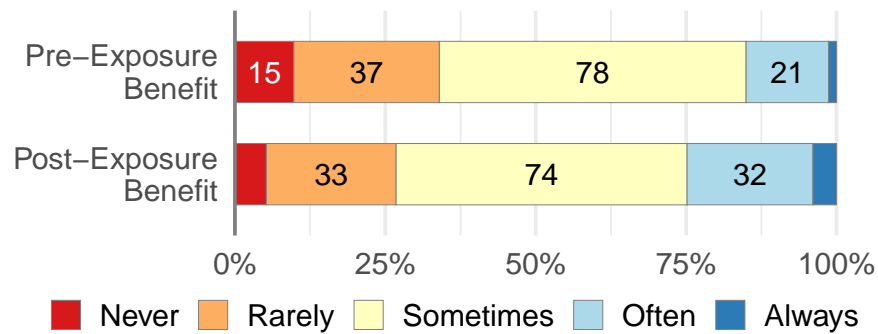
It shows me what I have done but not how they are using it or what they are collecting from this data. Like are they collecting what I do in the app, what I engage with, how long I’m there what keeps my interest. (P17)

How concerned are you with the amount of information Google is collecting about your activities online?



(a) Level of concern before and after visiting My Activity.

How often do you benefit from the amount of information that Google collects about your activities online?



(b) Frequency of benefit before and after visiting My Activity.

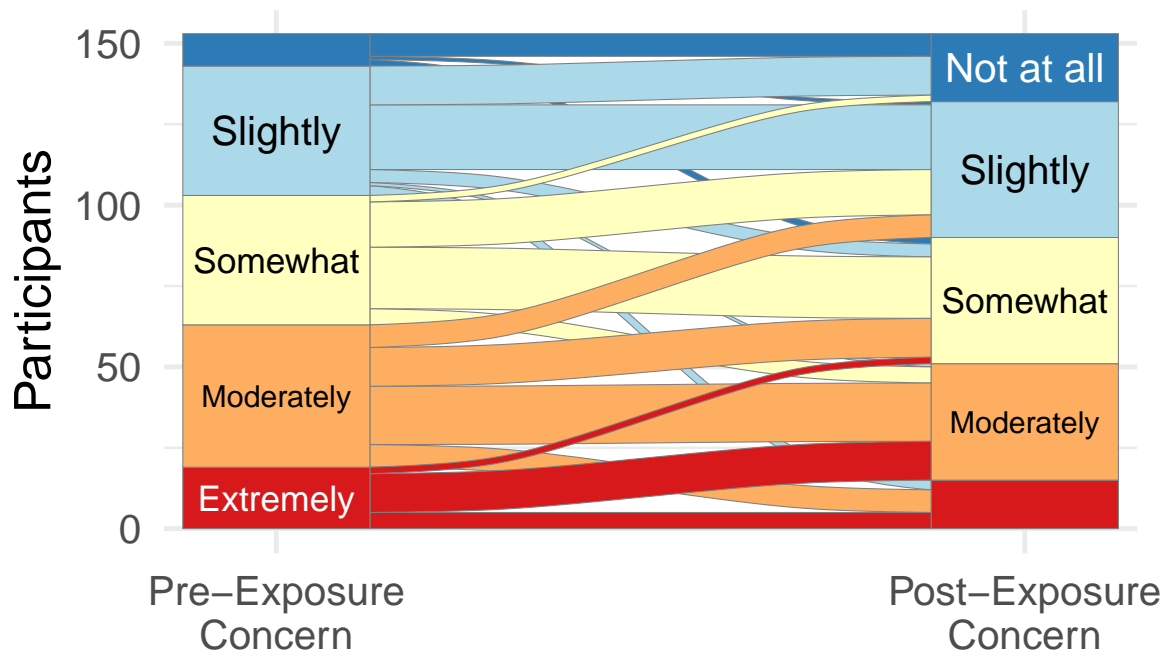
Figure 3.6: Proportions of the participants' assessment of (a) the level of concern (**Q2** & **Q26**) and (b) the frequency of benefit (**Q3** & **Q27**) before and after visiting the My Activity dashboard.

Purpose of Data Collection. We asked the participants to think of three purposes for which Google might collect this data (**Q7**). Most participants ($n = 123$; 80 %) stated that the purpose for the collection was targeted advertising. For example: “Make advertisements more targeted and effective” (P22), and “To target advertisements at me from my search history” (P29). The next largest group identified experience improvements that include personalization ($n = 109$; 71 %) as the purpose, e. g., “Customize my search results based on interest” (P39), and product improvements ($n = 42$; 25 %), e. g., “Usage data for company research for products and programs” (P149). Some participants ($n = 59$; 39 %) thought that Google’s purpose was to sell their usage data. P10 said, “Sell my data to third parties for profit,” and P31 said, “To sell to other companies.”

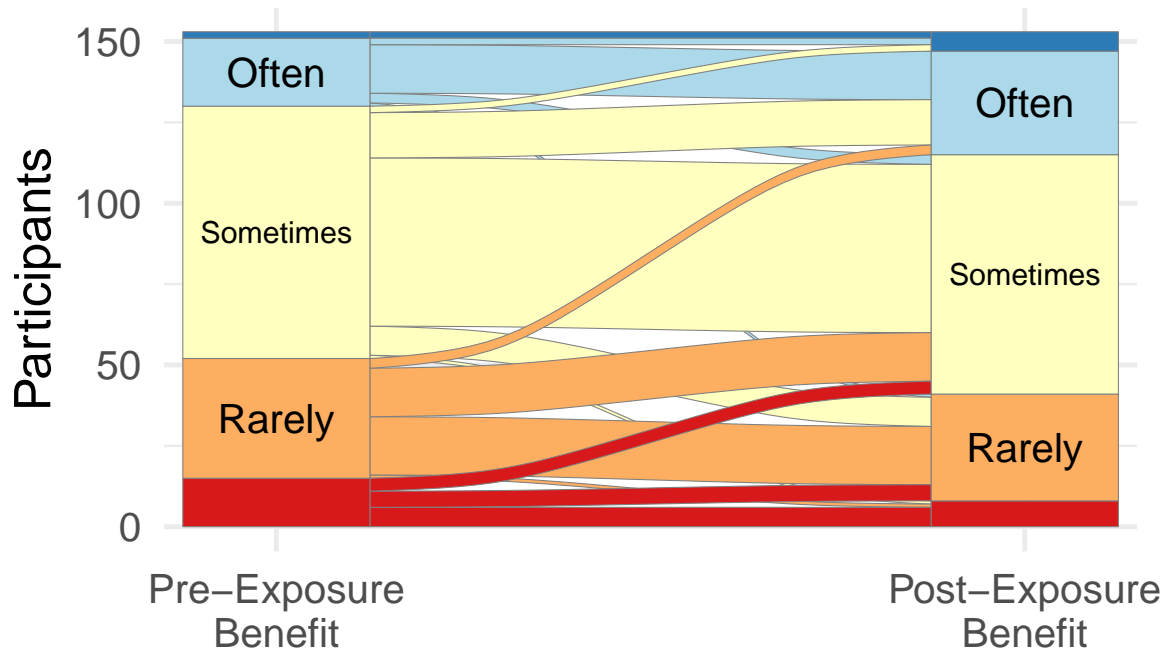
The purposes provided are mostly in line with what Google describes on its help pages, where they indicate the following reasons to collect activity data: (i) product improvements, (ii) recommendations, (iii) personalizations, and (iv) browser/search/location history. However, knowing the purpose for the data does not imply agreement with the use, and so we also presented participants with Google’s explanations for data collection, asking participants to gauge appropriateness of the explanation (**Q17-Q19**). For all three activity categories, Figure 3.5 shows that 64 % think the reasons to collect activity data are at least *slightly appropriate*.

3.3.2 RQ2: Impact on Benefit and Concern

Google’s My Activity dashboard provides extensive insights into data collection, and in this research question we seek to understand if exposure to My Activity affects concerns about or beliefs in benefits of Google’s data collection. We evaluate two Likert questions, one about concern (**Q2, Q24**) and one about benefit (**Q3, Q25**), before and after exposure to My Activity, as well as open-response explanations in answering this research question. The responses are visualized in Figure 3.6.



(a) Level of concern alluvium plot.



(b) Frequency of benefit alluvium plot.

Figure 3.7: Detailed visualization of how the participants change their assessments of (a) the level of concern (**Q2** & **Q26**) and (b) the frequency of benefit (**Q3** & **Q27**) after interacting with the My Activity dashboard.

Initial Perceptions Concerns. When participants were asked to explain their concern (Q2_A) with the amount of information Google is collecting, more than half of the participants ($n = 79$; 52 %) said they had privacy concerns, such as concerns about the amount of information ($n = 15$), e. g., “I feel like Google is taking way too much of my data” (P128), sensitivity of the information ($n = 14$), e. g., “I’m concerned that the data collected can be very specific and in turn, identifying” (P103), and feeling uncomfortable sharing information ($n = 12$), e. g., “My information is private and should be shared with no one” (P54). For some participants ($n = 29$; 19 %) the unknowns were concerning, such as how the information is used ($n = 19$), and who has access to the information collected ($n = 5$). For example P95 said, “I don’t know what is being done with my personal information that Google collects and who is capable of gaining access to it.” Security was also a concern for some participants ($n = 22$; 14 %), specifically concerns about data misuse ($n = 18$) and personal data being released ($n = 8$). This quote from P138 is an example: “I am concerned about any platform, application or website wrongfully accessing my data or having a breach of the data I provide.” Still others ($n = 16$; 11 %) responded that there existed a trade-off between privacy and free services, such as P115 who said: “I don’t like that my privacy is being compromised, but overall I enjoy the convenience of the services and feel its worth it.”

Initial Perceptions of Benefits. When explaining the benefit (Q3_A), participants described the benefits of improved suggestions ($n = 46$; 30 %), personalized advertisements ($n = 24$; 16 %), and the availability of usage history ($n = 15$; 10 %). For example, P11 said, “I’m given information and predictions about what I’m looking for in a more precise and efficient manner, because my data has clued Google in.” Participant P39 who found personalized advertisements useful said, “I receive ads that I have interest in and do not see ‘annoying’ ads as a result.” Participant P26 had this to say about the benefits data collection for usage history: “Use of My Activity helps me retrace my steps and find information that I may need at a later date.” Other participants ($n = 26$; 17 %) said they perceived no benefit,

such as participant P17, who said “All they do is bombard me with more ads and it doesn’t help me to do anything.”

Other Concerns. We additionally asked if participants had other concerns (Q8) prior to exposure, and many participants reported privacy concerns ($n = 58$; 38 %), security concerns ($n = 31$; 20 %), and too many unknowns ($n = 22$; 14 %). Among the privacy concerns were concerns about selling information ($n = 14$) and third parties ($n = 12$); for instance, participant 19 had this to say:

Google sells my information as a product. I am not really a customer. I am like a piece of corn that is sold on the commodities market. The farmer, Google, feeds my information and I respond. I am then sold to the highest bidder several times. (P19)

There were also privacy concerns about the amount of information ($n = 12$), e. g., “It’s just an odd feeling, knowing they collect every bit of information about me and keep it probably forever.” (P108)

Participants’ security concerns were about data breach ($n = 29$), e. g., “It does worry me if they ever had a data breach because it seems like they do have a lot of information about their users” (P143), and potential data misuse ($n = 18$) e. g., “I also worry about hacking and unsavory entities using my information in ways I don’t even understand” (P89), and “I am confident that given the opportunity, some human with any access to the data will use it for selfish reasons, possibly to the detriment of others” (P127).

Changing Level of Concern. To determine if there are significant changes in perceived concerns, we performed a Wilcoxon signed-rank test on the Likert responses before (Q3) and after (Q24) exposure to My Activity. We find that concern significantly decreases ($W = 2519.5$, $p < 0.001$) with an effect size of $r = 0.32$, suggesting that this decrease in concern is moderate in size.

To explore what factors may have influenced the decline in concern, we performed ordinal logistic regression with outcome variable of the Likert concern scale (see Table 3.2).

Table 3.2: Ordinal regression model to describe the level of concern after visiting the My Activity dashboard. The model uses a descending concern scale (i. e., from *extremely* to *not at all concerned*). The Aldrich-Nelson pseudo R^2 of the model is 0.63.

Factor	Est.	OR	Pr(> z)	
Pre-Exp. concern = <i>Extremely</i>	5.71	302.84	<0.001	***
Pre-Exp. concern = <i>Moderately</i>	4.56	95.60	<0.001	***
Pre-Exp. concern = <i>Somewhat</i>	2.77	15.96	0.002	**
Pre-Exp. concern = <i>Slightly</i>	1.18	3.26	0.167	
Increasing benefit	-0.17	0.85	0.654	
Knows My Activity = <i>Yes</i>	-0.32	0.72	0.348	
IUIPC cont. > 3.5	0.29	1.34	0.784	
IUIPC awar. > 3.5	-0.41	0.66	0.842	
IUIPC coll. > 3.5	0.29	1.33	0.595	
Gender = <i>Male</i>	-0.24	0.79	0.481	
Age $\in \{18 - 34, 25 - 34\}$	0.39	1.47	0.417	
Age $\in \{35 - 44, 45 - 54\}$	0.91	2.50	0.052	.
Edu. $\in \{No\ sch.g, (Sm.)\ HS\}$	0.19	1.21	0.734	
Edu. $\in \{Sm. col., Assoc., Prof.\}$	0.07	1.07	0.844	
Has IT background	0.64	1.90	0.105	
# of activities > median	-0.47	0.63	0.146	
Intercepts				
<i>Not at all</i> <i>Slightly</i>	0.58	1.79	0.773	
<i>Slightly</i> <i>Somewhat</i>	2.87	17.66	0.158	
<i>Somewhat</i> <i>Moderately</i>	4.63	102.0	0.024	*
<i>Moderately</i> <i>Extremely</i>	6.77	875.08	0.001	**
Signif. codes: *** $\hat{=}$ < 0.001; ** $\hat{=}$ < 0.01; * $\hat{=}$ < 0.05; . $\hat{=}$ < 0.1				

We included binary variables for initial concern, benefit increased, high IUIPC factors, gender, age, education level, IT background, and number of activities stored in the Google account, and the final model had a Aldrich-Nelson pseudo- $R^2 = 0.63$. We find that those who had *extremely* ($\eta = 5.71, OR = 303, p < 0.001$), *moderately* ($\eta = 4.56, OR = 96, p < 0.001$), and *somewhat* ($\eta = 2.77, OR = 16, p < 0.001$) concern initially were significantly likely to reduce their concern after exposure. Participants who were *extremely* concerned were $303\times$ more likely to reduce their concern, and those *moderately* concerned were $96\times$ more likely. All other factors seem to have no or little effect, except perhaps for the age range 35 to 54 ($\eta = 0.91, OR = 2.5, p = 0.052$).

Table 3.3: Ordinal regression model to describe the frequency of benefit after after visiting the My Activity dashboard. In the model a ascending frequency scale (i. e., from *never* to *always*) is used. The Aldrich-Nelson pseudo R^2 of the model is 0.68.

Factor	Est.	OR	Pr(> z)	
Pre-Exp. benefit = <i>Never</i>	22.29	4.81×10^9	<0.001	***
Pre-Exp. benefit = <i>Rarely</i>	20.26	6.28×10^8	<0.001	***
Pre-Exp. benefit = <i>Sometimes</i>	18.58	1.17×10^8	<0.001	***
Pre-Exp. benefit = <i>Often</i>	16.12	1.00×10^7	<0.001	***
Increasing concern	0.57	1.77	0.268	
Knows My Activity = <i>Yes</i>	-0.56	0.57	0.133	
IUIPC cont. > 3.5	0.00	1.00	0.998	
IUIPC awar. > 3.5	0.18	1.20	0.935	
IUIPC coll. > 3.5	0.41	1.50	0.473	
Gender = <i>Male</i>	0.92	2.51	0.014	*
Age $\in \{18 - 34, 25 - 34\}$	0.76	2.14	0.126	
Age $\in \{35 - 44, 45 - 54\}$	0.58	1.79	0.256	
Edu. $\in \{No\ sch.g, (Sm.)\ HS\}$	-0.16	0.85	0.782	
Edu. $\in \{Sm. col., Assoc., Prof.\}$	-0.16	0.85	0.671	
Has IT background	-0.07	0.93	0.864	
# of activities > median	0.30	1.36	0.373	
Intercepts				
<i>Always</i> <i>Often</i>	15.61	6.00×10^6	<0.001	***
<i>Often</i> <i>Sometimes</i>	18.78	1.43×10^8	<0.001	***
<i>Sometimes</i> <i>Rarely</i>	21.99	3.55×10^9	<0.001	***
<i>Rarely</i> <i>Never</i>	24.56	4.63×10^{10}	<0.001	***
Signif. codes: *** $\hat{=}$ < 0.001; ** $\hat{=}$ < 0.01; * $\hat{=}$ < 0.05; · $\hat{=}$ < 0.1				

The alluvium plot in Figure 3.7a shows in more detail how the level of concern changes among the participants based on their initial concern. In total, 61 (40 %) participants moved down the scale, 69 (45 %) stayed the same, and only 23 (15 %) increased their concern.

Changing Perceptions of Benefits. We find that there is a significant increase in perceived benefit (Wilcoxon signed-rank test, $W = 435$, $p < 0.001$) with a moderate effect ($r = 0.32$). Using the same factors as before, we constructed an ordinal logistic regression model to identify potential covariants that led to the increase in benefit (see Table 3.3).

Across all initial benefit responses (*never*, *rarely*, *sometimes*, *often*, and *always*), the

regression exposes significant likelihood of keeping the same benefit or increasing benefits with odds ratio $> 10^6$, suggesting that participants across the spectrum recognized benefits to Google's data collection. We observed that participants identified as male also were significantly more likely to increase their benefit perceptions ($\eta = 0.92, OR = 2.5, p = 0.014$), but other factors were not meaningfully significant.

Figure 3.7b provides more insights into the broad increase in perceived benefit. In total, 45 (29 %) increased their benefit response, 93 (61 %) kept it the same, and only 15 (10 %) decreased their perceived benefit.

Final Perceptions of Concerns. Post exposure, participants were also asked to explain their final concern (Q26_A) choices. Qualitative coding revealed that while the total number of participants describing a privacy concern dropped from 79 (52 %) to 72 (47 %). The number of participants who described privacy concerns about the amount of information collected increased to 21 (14 %) from 15 (10 %). For example, P22 said:

I'd say I'm a little more concerned now about how much is being collected. Especially with one of the random activities shown in the survey being well over a year old.

Similarly, some participants, 25 (16 %) versus 22 (14 %), mentioned security concerns, and prevalent codes that increased included data misuse ($n = 20$), e. g., "I'm worried about the misuse of the data and security of it" (P127), and personal data being released ($n = 14$), e. g., "There is always a chance that personally identifiable information can somehow be leaked to the Internet at large" (P147).

We also observed a slight increased in participants describing that they were now unconcerned with the data collection, 25 (16 %) versus 21 (14 %). For example, participant P13 said:

I am not concerned. Nothing's ever gone wrong as a result of what they collect. I don't have things to hide. I imagine the data collection helps me.

Final Perceptions of Benefits. Explaining their final benefits from Google’s data collection (**Q27_A**), qualitative coding revealed an increase in the number of participants who described benefits of suggestions: 70 (46 %) versus 46 (30 %). For example participant P20 said, “YouTube recommendations are tailored around my activity, so that’s beneficial to me,” and participant P119 said, “Many of Google’s services offer useful personalized suggestions based on my data.” We also found an increase in the number of participants who said that they found that access to their activity history beneficial: 26 (17 %) versus 15 (10 %, e. g., “I can go back to a websites I have viewed about specific things if I need to, find that song I really want to hear that I listened to last week, and make it easier to get places I may be returning to” (P18). Fewer participants said they received no benefit from the data collection: 19 (12 %) versus 26 (17 %), like P86 who said, “How and why would I benefit from it when I didn’t even know they are collecting information about my activities.”

3.3.3 RQ3: Behavioral Change

To answer our third research question **RQ3**, we surveyed participants about their willingness to take action after they have learned about Google’s data collection practices. We asked three closed-ended (**Q23**, **Q24**, and **Q25**) and three open-ended questions (**Q23_A**, **Q24_A**, and **Q25_A**) to gauge participants’ intentions to take action or change their behavior as a result of the exposure to My Activity. The results of the three closed-ended questions are summarized in Figure 3.8.

Change Account Settings. We asked the participants to indicate whether they plan to adjust some of the (privacy) settings after seeing their My Activity page (**Q23**). The results were almost the same for *yes* ($n = 57$; 37 %) and *no* ($n = 56$; 37 %) while 26 % ($n = 40$) of the participants were *unsure*.

We constructed a logistic regression model to identify factors that predict the outcome of being willing to change settings. We included covariants for change in concern, change

in benefit, IUIPC responses, demographics, and total number of activities over the lifetime of an account. The model did not expose any significant factors.

We also qualitatively explored participants views about their privacy settings by asking them which settings, if any, would they change (**Q23_A**). More participants ($n = 75$; 49 %) in their qualitative answer responded with a privacy setting that they would change. We recognize that many of these participants do not plan to actually change settings given their prior quantitative responses. We found that changing delete frequency ($n = 16$), or stopping data collection ($n = 17$), or changing how long information is stored for specific things ($n = 27$) are the most popular reasons to revise the settings. For example P22 said, “I would update when they delete my data so it stays current, relevant, and up-to-date.” P75 said, “I would have my settings changed so that it no longer stores any data,” and P120 said, “Probably auto delete since I don’t remember to go delete it often enough.”

Other participants ($n = 52$; 34 %) reported that they would not change their settings. One reason for not changing the settings was that the participant likes the current settings. For instance P13 said, “I have no complaints so see no reason to fix something that isn’t broken.” Another reason was that they had already configured the settings. For example P133 said, “I’ve already used this page and configured it the way I want.”

Participants who were undecided ($n = 21$; 14 %) about changing the settings stated they wanted to review the settings. Like participant P45 who said, “I need to look more into the settings to see something I may change.” Undecided participants also reported that they wanted to review the data collection, e. g., “I’d at least want to actually take a look and see just how much is collected, with using my account across all devices, and how far back my activity goes” (P22).

Review or Delete Activities. When asked whether they plan to use My Activity again after the survey for reviewing or deleting activities (**Q24**), only 37 % of the participants responded with *yes*. The remaining 63 % ($n = 56$) were either *unsure* ($n = 48$; 31 %) or said

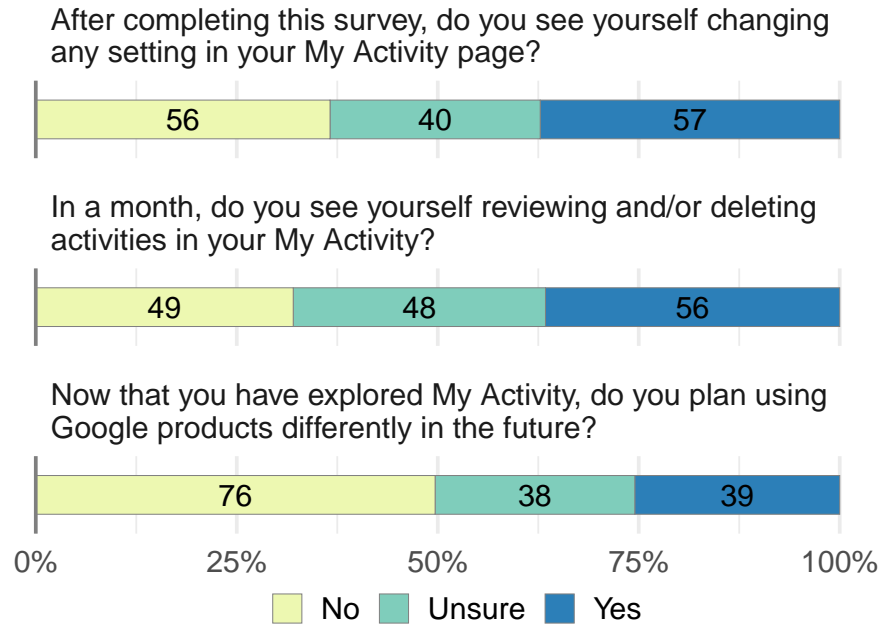


Figure 3.8: Willingness to take action after visiting the My Activity dashboard.

no ($n = 49$; 32 %).

We performed logistic regression to determine factors that would lead to reviewing activities (see Table 3.4). We found a significant correlation with IUIPC collection scale questions ($\beta = 1.43$, $OR = 4.19$, $p = 0.042$), where participants with high privacy concerns regarding data collection were $4.19\times$ more likely to review activities later. This finding suggests that individuals predisposed to have concerns about data collection are likely to benefit the most from My Activity.

We also qualitatively coded participants' explanations for why they would or would not review their activities (Q24_A). The main reason participants gave for continuing to use My Activity was to delete activities ($n = 90$; 59 %). The most common activities participants said they would delete were Search ($n = 31$), Maps ($n = 23$), and YouTube ($n = 19$). For others ($n = 12$) it was activities of a sensitive nature that they would return to delete. For example, participant P89 said:

Personal activities. Like I noticed that there were medical searches in my activities. It makes me uncomfortable that information is taken about me.

Table 3.4: Binomial logistic model to describe which other factors (beside visiting My Activity) influenced the participants plan to review/delete activities (*yes* responses to Question Q24). The Aldrich-Nelson pseudo $R^2 = 0.25$.

Factor	Est.	OR	Pr(> z)
(Intercept)	13.77	9.53×10^5	0.988
Increasing concern	0.90	2.45	0.068 .
Increasing benefit	0.05	1.05	0.900
IUIPC cont. > 3.5	0.63	1.88	0.613
IUIPC awar. > 3.5	-16.69	5.63×10^{-8}	0.985
IUIPC coll. > 3.5	1.43	4.19	0.042 *
Gender = <i>Male</i>	0.45	1.56	0.255
Age $\in \{18 - 34, 25 - 34\}$	-0.07	9.32×10^{-1}	0.893
Age $\in \{35 - 44, 45 - 54\}$	0.24	1.27	0.652
Edu. $\in \{No\ sch.g, (Sm.) HS\}$	0.85	2.34	0.179
Edu. $\in \{Sm. col., Assoc., Prof.\}$	-0.04	9.64×10^{-1}	0.928
Has IT background	0.34	1.40	0.441
# of activities > median	-0.35	7.02×10^{-1}	0.330
Signif. codes: *** $\hat{=}$ < 0.001; ** $\hat{=}$ < 0.01; * $\hat{=}$ < 0.05; . $\hat{=}$ < 0.1			

Other participants ($n = 41$; 27 %) reported that they do not plan to use My Activity in the future. Reasons included making changes would be too time consuming ($n = 7$), e. g., “I have better things to do with my time, frankly, than to be reviewing this” (P92), or that they would easily forget to do so ($n = 3$), e. g., “Honestly, I’ll probably forget about it, so I’m unlikely to delete things a month from now” (P129).

Still others were ($n = 12$; 8 %) undecided. For instance participant P36 said, “I’m not sure, I would have to weigh convenience for me vs. the feeling of too much information being collected.”

Use Google Differently. Nearly 50 % ($n = 76$) of the participants did not plan to use Google products differently in the future in response to Q24. The remaining responses split evenly between *yes* ($n = 39$; 26 %) and *unsure* ($n = 38$; 25 %).

We performed a logistic regression to determine factors that may influence reported changes in behavior (see Table 3.5). Unsurprisingly, we found two significant factors. Those

Table 3.5: Binomial logistic model to describe which other factors influenced the participants plan to use Google products differently in the future (*yes* responses to Question **Q25**). The Aldrich-Nelson pseudo $R^2 = 0.42$.

Factor	Est.	OR	Pr(> z)	
(Intercept)	-16.23	8.93×10^{-8}	0.997	
Increasing concern	2.50	1.22×10^1	<0.001	***
Increasing benefit	-1.31	2.71×10^{-1}	0.027	*
IUIPC cont. > 3.5	18.42	1.00×10^8	0.989	
IUIPC awar. > 3.5	-4.82	8.06×10^{-3}	0.999	
IUIPC coll. > 3.5	1.40	4.07	0.109	
Gender = <i>Male</i>	-0.73	4.80×10^{-1}	0.146	
Age $\in \{18 - 34, 25 - 34\}$	0.97	2.63	0.198	
Age $\in \{35 - 44, 45 - 54\}$	1.21	3.36	0.109	
Edu. $\in \{No\ sch.g, (sm)\ HS\}$	0.45	1.57	0.562	
Edu. $\in \{Sm\ col.\ Assoc., Prof.\}$	-0.16	8.56×10^{-1}	0.757	
Has IT background	0.86	2.37	0.114	
# of activities > median	-1.25	2.86×10^{-1}	0.006	**
Signif. codes: *** $\hat{=}$ < 0.001; ** $\hat{=}$ < 0.01; * $\hat{=}$ < 0.05; . $\hat{=}$ < 0.1				

who had an increase in concern ($\beta = 2.50, OR = 12.2, p < 0.001$) and a decrease in (or same) benefit ($\beta = -1.31, OR = 0.27, p = 0.027$) were significantly more likely to use Google products differently. This represents a small fraction of participants in our study: 23 (15 %) participants noted an increase in concern, 15 (10 %) reported a decrease in benefit.

In addition, we found a third significant factor. Participants whose accounts contained a high number of activities (i. e., more than the median number of activities) were significantly more likely to report to use Google products differently ($\beta = -1.25, OR = 0.29, p = 0.006$).

Looking at the qualitative results shows that of those who planned to use Google products differently some would change settings ($n = 14; 9\%$), such as limiting data collection or deleting their activities more often, e. g., “I am definitely going to be turning off history for YouTube while working” (P147). Others would change the way they use Google products and services more generally ($n = 12; 8\%$), such as being more careful when using them, e. g., “I’d certainly be aware of what was being collected and modify my searches accordingly” (P148). Some participants would start to limit their usage of Google products and services

($n = 9$; 6 %), e. g., “I would use less of Google and more of other services” (P96).

Of those participants who were unsure if they would change using Google products, some ($n = 10$; 7 %) stated that change would be difficult because of the importance of Google products; for instance P6 said:

I realize Google products are necessary to my lifestyle and work, but I also like to be in control of my data. I'm not sure what the best course of action is at this point.

Of the participants who would not change the way they use Google products, many ($n = 37$; 24 %) claimed they were happy with the status quo, like participant P139 who said, “I am happy with the current setup and will continue as I always have.” Some participants ($n = 25$; 16 %) were simply unconcerned, e. g., “I just don’t care enough from what I saw to change how I use Google” (P122). Others ($n = 6$; 4 %) are simply privacy resigned, e. g., “I’ve accepted the fact that they work this way whether I view it as right or not” (P120).

Willingness to Pay for Google’s Services. We asked participants if they were willing to pay for Google services if activity data were not collected (**Q16**), and those results are presented in Figure 3.9. Nearly half of the respondents ($n = 74$; 48 %) would not pay, which is in line with previous work [22], but 70 (46 %) say they would pay at least \$1 USD per-month, with large clusters at \$10 USD, \$5 USD, and \$1 USD per-month. Only 10 (7 %) described a willingness to pay more than \$10 USD per month.

The average revenue per user (ARPU) is currently not reported by Google and differs significantly between regions. According to eMarketer [23] Google will make a *net ad revenue* (after paying traffic acquisition costs to partner sites) in the U.S. of approximately \$39.58 billion USD in 2020. As of December 2020, Google had close to 271 million unique monthly visitors in the U.S. [27], resulting in an ARPU of \sim \$146 USD (Facebook \$159 USD [43]), or roughly \$12 USD per month. This is in line with Google’s pricing for workspace accounts (\$12 USD per-user and per-month), and thus, one can assume that Google would require a fee of \geq \$12 USD per month (but likely more) in return for not

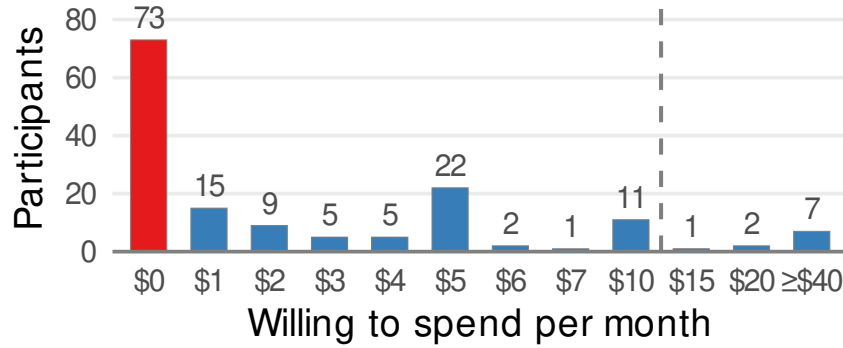


Figure 3.9: More than half of the participants ($n = 80$; 52 %) stated to pay at least \$1 USD per month for Google services. These are opposed by 73 (48 %) who would not pay any money. Only 10 participants signalized a willingness to pay more than \$12 USD per month (indicated by the dashed line).

collecting data. Our data suggests that only a small fraction would pay such a fee, and perhaps fewer, as this result could be affected by response bias.

3.4 Discussion

Controlling Data Collection. Participants reported significantly higher benefits from and lower concerns about Google’s data collection after interacting with My Activity. These shifts could be accounted for by an increased awareness of the collection and the prospects of being in control of that collection with access to the history of activities, similar to what Schnorf et al. found [132]. This may especially be the case in our study as only $\sim 50\%$ of respondents reported being aware of specific activity displayed during the survey, and $\sim 75\%$ agreed or strongly-agreed that My Activity helps them to better understand Google’s data collection.

The notion of information flow controls is an important factor in privacy perception [140]. Interactions with My Activity increase the subjective (and also objective) control over collected data, reducing concern in relation to an original feeling of lack of control and an inability to restrict data access. My Activity and other data collection transparency and management tools are both in the end-users’ and service providers’ best interest, and we

expect (and hope) that more online services will provide such tools in the future.

Opaque Control Choices. My Activity allows users a plethora of choices, but it may be too difficult for users to make informed decisions about individual activities as the impact of keeping or deleting individual items is opaque to the user. In their study of ad profiles, Rao et al. suggested that the dashboards did not provide transparency on how or why user profiles were created [117], and this lack of additional information inhibits clear decision making. The only current explanations on My Activity suggest that the experience will degrade, but specifically why deleting any given activity or bundle of activities degrades experiences (or improves privacy) is not readily available.

There is evidence that providing some more insight into inferences could be beneficial, as users tend to relate inferences with their past activities [116], and there already exists language from Google that suggests the seasonality of data matters in inferences; this motivates the deletion time frames of 3- and 18-months [112]. Expanding the options for how to manage data collection, perhaps based on inferences made or other metrics, would better assist users in making clearer choices in managing their activity data.

Management at Scale. Services like My Activity put significant pressure on users to continuously and individually manage their data collection, especially, as new data collection occurs all the time, and in some cases, users may have to review activities across multiple accounts. It is likely that well intentioned data-privacy laws, like GDPR, may lead to increased data collection management due to data access requirements. The truth is, such management does not scale, and the benefits of increased control could be neutered by the increase in the scale of decision making.

We see evidence for My Activity that users are unlikely to take advantage of these controls, perhaps due to the scale of the management requirement not just with Google but elsewhere. This is only one of possibly many transparency tools; mechanisms for secure and transparent umbrella management of data collection across services is likely to be needed

as privacy dashboards and data-rights laws proliferate. Such umbrella services have been proposed previously in the literature [6, 51, 119, 152], and with apps like *Jumbo* [101, 105], some first real-world tools exist. However, these umbrella services need to find the balance between displaying relevant information to the users but not overwhelming them. Exploring whether and how these services could reduce management at scale is an area of future research.

Lack of Negative Consequences. Only about a quarter of respondents indicated they would change their behavior as a result of increased awareness of Google’s data collection, and this sentiment is entirely understandable given how integral Google’s products and services are in the online experience. A number of participants explicitly noted that they trust Google, and thus after exposure to their My Activity page remained unconcerned about Google’s data collection. As there is no information of potential privacy risks on the My Activity page, it is not surprising that people are unconcerned with and not aware of how their data can be used in expected ways by either Google or third-parties.

For those who indicated they would change settings, this group reported high on the UIPC Control scale, indicating that individuals who seek more control over their data will likely take advantage of such a service. The remaining users are less likely to do so, perhaps because they have not experienced negative consequence and instead rely on the default policy, which may not be in their best interest. Increased exposure and encouragement for users to understand the benefits and risks of data collection could lead to better outcomes for everyone, as it may encourage service providers to use better default privacy settings.

Design Implications. Based on the findings of our study, we offer some suggestions to improve the utility of privacy dashboards. 1) Provide concrete explanations for which purpose activities are collected and stored. For instance, when activities are used to infer interests of a person, make this link between the activity and inference more explicit (e. g., search query for “Seattle Seahawks” results in the inference “American Football” and

the aggregated inference “Sports”). 2) Participants felt overwhelmed with the amount of activities being collected and presented to them. It is worthwhile to explore ways to give users a better overview of and means to navigate through their activities. Showing simple statistics (e. g., the number of activities grouped by month or service) might helping people to better grasp the amount of activities collected. Activities could also be further clustered beyond the existing My Activity bundle view, which groups by time and Google product. For example, each cluster could be further grouped by broader themes, e.g., by inferences applied for advertising, that could assists users to better focus on activities that may need manual review. 3) Some participants expressed the need for better ways to remove certain activity classes, for example, any search related to medical issues. Offering keyword management strategies where users can custom define activity deletion policies based on user defined criteria would help users manage their privacy without having to regularly inspect their activities. In May 2021, after data collection, Google introduced a “quick delete” feature which removes the last 15 minutes of search activities [52].

However, adding too much functionality carries the risk of overwhelming users with a complex UI, discouraging its use. The simplicity of My Activity’s design is admirable, but this needs to be balanced with providing substantive information about the purpose the data is collected and how it will be used. Designing a more effective transparency tool that is both simple and deeply informative requires more exploration.

3.5 Conclusion

In this work, we sought to understand how privacy dashboards and transparency tools affect concerns about and benefits from data collection. Focusing on Google’s My Activity tool, we conducted a pre-post-study where participants answered questions about concern/benefit before and after exposure to My Activity. We find that My Activity significantly decreases concern about Google’s data collection practices and increases the perceived benefit, despite participants qualitatively stating the same concerns and benefits before and after exposure.

Transparency tools, like My Activity, are clearly beneficial to the service providers and can also support data management for the user. We, unfortunately, find that most participants are unwilling or unsure if they will review their activities following this study.

Chapter 4: Security and Privacy Perceptions of Third-Party Application Access for Google Accounts

The 2018 Cambridge Analytica scandal [77] prompted the scrutiny of third-party apps that integrate with online application programming interfaces (APIs). This came about when an online personality quiz used the Facebook API to collect detailed personal information from millions of unsuspecting quiz-takers and their friends. In response, Facebook restricted how third-parties can use its API [133]. However, Facebook is not unique among online services that allow third-party apps to leverage user account data via APIs.

Likewise, Google has APIs that allow third-party services to use existing user account data. For example, Google's single sign-on (SSO) [64] lets users log onto participating third-party services with their already-existing Google credentials. Other major online services like Apple [10], Twitter [142], and Facebook [100] offer similar SSO capabilities using their accounts.

The Google API also exposes functionality from various Google products. For instance, APIs let authenticated third-party apps to interact with Google Calendar entries or Gmail correspondence on behalf of the user. This particular integration is what enables iOS's built-in Calendar app to display and edit a user's Google Calendar events, among others.

Despite many benefits, granting programmatic access to one's online account can pose security and privacy risks, as highlighted by Cambridge Analytica. In 2018, Google proposed granular permissions for API authorizations to give users more control and mitigate these risks [62]. As of September 2021, however, this updated design does not appear to be widely adopted. Google API integrations in popular services like Dropbox and Zoom still use all-or-nothing consent flows, and a spot check of the most popular apps on the Google Workspace Marketplace¹ shows those too lack fine-grained permissions.

¹<https://workspace.google.com/marketplace/category/popular-apps>

In this paper we explore how users consider security and privacy in light of third-party API access to their Google accounts given the disclosure and control mechanisms currently available. First, we surveyed $n = 432$ participants to recall the last times they used Google SSO or authorized a third-party app access to their Google account data. When recalling these, we also asked participants what factors they considered before granting access. Of the 432 participants, the vast majority (89%) recalled using SSO, but only half (52%) recalled granting the third-party access to their Google account.

We then invited $n = 214$ participants from the first survey to return for a follow-up survey. As part of this second survey, participants installed a browser extension that parsed entries in their Google account’s “Apps with access to your account” dashboard.² Based on this data, we asked participants about specific apps they currently have installed on their Google account. From the browser extension, we observed 1,010 third-party services that use Google SSO and 455 third-party apps that integrate with APIs for various Google services. Of the observed third-party apps, nearly half require two or more permissions accessing the participants’ Google account. The most common permission is modifying Google Play Game activity (223 instances), followed by viewing primary Google email address (189), and viewing personal info (177).

Participants were overall only *Slightly concerned* or *Not concerned* about the access granted to third-party apps, but showed the most concern about apps viewing personal info; 39% were *Very concerned*, *Concerned*, or *Moderately concerned*. Interestingly, such information is perhaps less of a privacy and security risk than third-party apps that can modify/view contacts, email, calendar events, or cloud storage. The relative lack of concern with these permissions could be attributed to a transference of trust to Google, as evidenced by open-ended responses where participants indicated that they believe Google is properly vetting these accesses.

We surveyed participants about the specific apps on their accounts and asked if they

²<https://myaccount.google.com/permissions>

wished to keep or remove account access for those apps. A logistic regression revealed that participants were $5.8\times$ more likely to want to remove access for an app when they wished to change which Google account data the app can access. Additionally, they were $5.9\times$ more likely to keep access when the app was recently used, and $6.0\times$ more likely to keep access when they viewed the app as beneficial. However, 79% and 78% of participants indicated that they currently *Rarely* or *Never* review their apps and SSOs, respectively. After viewing their third-party accesses as part of our survey though, the vast majority (95%) of participants indicated they would want reminders to review those at least *Once a year*.

These findings suggest a significant opportunity to improve how users interact with third parties with programmatic access to their accounts by helping users to identify and remove less frequently used apps/SSO in an automated way, or to simply revoke access after a period of disuse. Similarly, Google could require regular re-approval of access, perhaps yearly so as not to be too disruptive. Additionally, many participants articulated a desire for controls of the permissions for third-party apps. This would allow users to better limit which aspects of their Google account each app/SSO can access with respect to the benefit being provided, rather than forcing them to accept an all-or-nothing approach.

4.1 Method

We begin by describing the two surveys. The first survey asked participants to recall prior experiences with third-party apps and SSOs. The second survey leveraged a custom browser extension and asked participants to respond to the specific SSOs and third-party apps currently authorized on their Google account. In the remainder of this section, we detail our study procedures, describe how we recruited participants, discuss ethical considerations of our study, and outline the limitations of our approach.

First Survey. Below, the first survey is described. The full survey can be found in B.1.1.

1. Informed Consent: Participants consented to the study.

2. Google Account Use: Participants were asked if they have a Gmail account (as surrogate for a Google account), if it is their primary Google account with sole ownership, and the age of the account. Questions: **Q₁1–Q₁4**.
3. Familiarity with “Sign in with Google”: Participants were provided with contextual information from Google’s documentation [64] alongside a screenshot of a “Sign in with Google” button (taken from Yelp) and asked about their recent experiences using their Google account to sign into a third-party app or service. Questions: **Q₁6–Q₁9**.
4. Familiarity with Third-Party App Account Access: Participants were provided with contextual information from Google’s documentation [61] alongside a screenshot of a third-party app’s Google API authorization screen (taken from a generic app). Next, participants were asked about their recent experiences granting a third-party app access to their Google account. Questions: **Q₁10–Q₁13**.
5. IUIPC-8: Participants answered the Internet users’ information privacy concerns (IUIPC-8) questionnaire [67], to gain insights into participants’ privacy concerns.
6. Demographics: Participants were asked to provide demographic information, such as age, identified gender, education, and technical background. Questions: **D1–D4**.

Second Survey. The second survey recruited from those who completed the first survey. We used the following inclusion criteria to ensure participants have active Google accounts with SSOs and/or third-party apps: (i) the participant has a Gmail account, (ii) the participant uses the Gmail account as their primary Google account, (iii) the participant has sole ownership of their Google account, (iv) the participant has used their Google account for more than one month, (v) the participant correctly answered the attention checks. Below, we describe each part of the second survey in detail, and the full survey can be found in B.1.2.

1. Informed Consent: Participants consented to the main study, which included notice that they would be asked to install a web browser extension that would access their Google’s “Apps with access to your account” page.

2. Extension Installation: Participants installed the browser extension that locally parses third-party apps and SSOs and displays specific apps to the participant as part of the survey. The extension also recorded aggregate information about the number of SSO and API authorizations and the date of the oldest and newest authorization.
3. Explore Apps with Access to Your Account: We provided participants with a brief descriptive introduction and then directed them to explore their Google “Apps with access to your account” page for one minute. This interaction was managed by the browser extension with an overlay banner and restricted navigation away from the page.
4. Account Access Questions: Participants were asked what they consider before allowing third parties access to their Google account and services and how often that access is reviewed. Questions: **Q₂1–Q₂4**.
5. Keep or Remove: Each participants was shown all their Google account authorizations and if they wished to keep or remove the authorization. Question: **Q₂5**
6. App Permissions: Participants were asked about the permissions for their newest, oldest, and a random third-party app to investigate potential impacts of installation time on concern, benefit, and recall. Questions: **Q₂7–Q₂14**.
7. Reflections: Participants were asked to reflect on their understanding of the Google “Apps with access to your account” page and if they would change their behavior as a result of that interaction. Questions **Q₂15– Q₂20**.
8. Feature Improvements: Participants provided suggestions for improving Google’s “Apps with access to your account” page. Questions **Q₂21–Q₂26**.
9. Uninstall Extension: Upon completing the survey, participants were instructed to remove the browser extension.

Recruitment and Demographics. We recruited 432 participants via *Prolific*³ for the first survey between March 31, 2021 and April 20, 2021. After applying our inclusion criteria, 399 participants qualified for the second study, and 214 returned to complete the

³<https://www.prolific.co> - Prolific participant recruitment service, as of October 19, 2022.

Table 4.1: Demographic and IUIPC data collected at the end of the first survey.

		First Survey (<i>n</i> = 432)		Second Survey (<i>n</i> = 214)	
		n	%	n	%
Gender	Woman	204	47	99	46
	Man	216	50	107	50
	Non-binary	11	3	7	3
	No answer	1	0	1	1
Age	18–24	132	31	66	31
	25–34	157	36	80	37
	35–44	75	17	32	15
	45–54	44	10	20	9
	55+	23	6	15	7
	No answer	1	0	1	1
IUIPC		Avg.	SD	Avg.	SD
	Control	5.9	1.0	5.8	1.0
	Awareness	6.6	0.7	6.6	0.8
	Collection	5.4	1.2	5.3	1.2
	IUIPC Combined	5.8	0.8	5.8	0.8

second survey. Participants received \$1.00 USD and \$3.00 USD for completing the first and second survey, respectively, and it took, on average, 8 minutes and 13 minutes to complete, respectively.

Thirty-one percent of first survey participants were between 18–24 years old, 36 % were between 25–34 years old, and 33 % were 35 years or older. The identified gender distribution for the first survey was 50 % men, 47 % women, and 3 % non-binary, self-described, or choose not disclose gender.

Thirty-one percent of second survey participants were between 18–24 years old, 37 % were between 25–34 years old, 31 % were 35 years or older, and 1 % chose not to disclose their age. The identified gender distribution for the second survey was 50 % men, 46 % women, and 4 % non-binary, self-described, or choose not disclose gender. Participant characteristics are presented in Table 4.1.

Additional demographic information can be found in Appendix B.2.

Analysis methods. When performing quantitative analysis, descriptive or statistical tests, the analysis is provided in context. For qualitative responses, we use open coding to

analyze responses to open-text questions. First, a primary coder from the research team crafted a codebook and identified descriptive themes for each question. A secondary coder coded a 20 % sub-sample as a consistency check, providing feedback on the codebook and iterating with the primary coder until inter-coder agreement was reached (Cohen’s $\kappa > 0.7$, $mean = 0.81$, $sd = 0.05$).

Ethical Considerations. The study protocol was approved by The George Washington University Institutional Review Board (IRB) with approval number NCR202914, and throughout the process we considered the sensitivity of participants’ Google app authorization data at every step. All aspects of the survey requiring access to the actual Google “Apps with access to your account” page was administered *locally* on the participant’s machine using the browser extension. All participants were informed about the nature of the study prior to participating and consented to participating in both surveys. At no time did the extension or the researchers have access the participants’ Google password or to any other Google account data, and all collected data is associated with random identifiers.

Limitations. As an online survey, we are limited in that we cannot probe deeply with follow-up questions to understand the full range of responses. We attempt to compensate for this limitation by performing thematic coding across many responses to capture general opinions and feelings when interacting with third-party apps and SSOs that have access to their Google account.

Additionally, we are limited in our recruitment sample, which is generally younger than the population as a whole. Yet, we argue that despite this limitation, our results offer insights into user awareness of third-party apps and SSO access to their Google account, as well as other online services with third-party APIs. We note that prior work by Redmiles, et al. [121] suggests online studies about privacy and security behavior can still approximate behaviors of populations.

Some results may be affected by social desirability bias, where participants might

indicate behavior that they believe we (the researchers) expect them to embody. For example, this may lead to participants over describing their awareness or recall of granting access to third-party apps or SSOs, or their intention to remove access. In these cases, one may view these results as a potential upper bound on true behavior.

Finally, we acknowledge our study only considers Google, even though many online services offer APIs. Still, we believe our findings are broadly applicable because of the consistency of the underlying mechanisms (i.e. OAuth scopes) and user consent flows (i.e., users grant install-time permissions with a dashboard for review) across many major providers.

4.2 Results

In this section we present the results of the two surveys. We first describe the observed apps, SSOs, and their permissions for participants completing the second survey. Next, we explore participants' awareness and understanding of third-party apps and their permissions. We then offer results on participants' motivations to install or not install a third-party app and their intentions to change settings. Finally, we report on participants' reflections on their third-party apps and desired features for improving transparency and control over API access to their Google account.

4.2.1 Measurements

Third-party app and SSO findings. In the first survey, 432 participants self reported if they recalled using Google's SSO (**Q16**) or granting third-party apps (**Q110**) access to their Google account. We found that 89% ($n = 386$) of participants recalled using SSO to sign into a third-party service. Furthermore, 52% ($n = 225$) recalled granting a third-party app access to their Google accounts. See Figure 4.1.

Additionally, during the first survey, we asked participants to recall the latest app or service they signed into using their Google account (**Q17**), and classified their responses

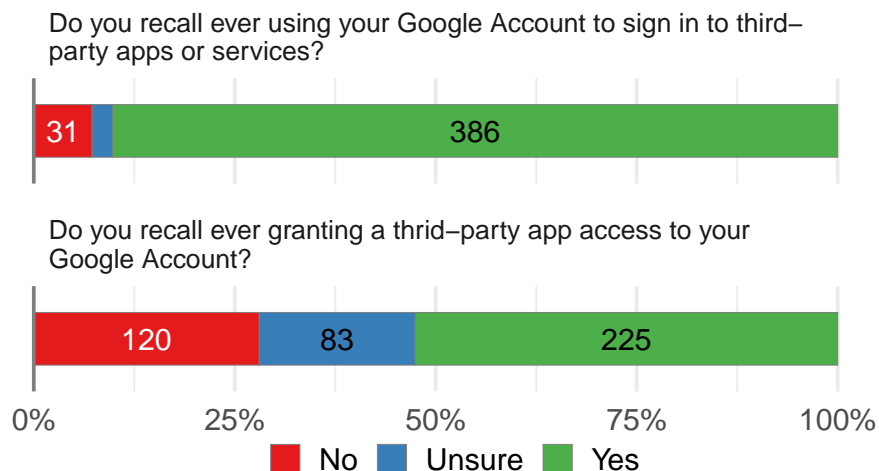


Figure 4.1: Ninety percent of participants recall using their Google Account to sign in (Q₁₆) and over half recall granting a third-party app access to their Google Account (Q₁₀).

into various categories based on the app or service. We searched repositories of available third-party apps, e.g., Google Play or Google Workspace Marketplace, to determine the proper app category. The top categories include shopping ($n = 51$; 12%), social media ($n = 42$; 10%), gaming ($n = 38$; 9%), food ($n = 34$; 8%) and entertainment ($n = 28$; 6%).

Among the 214 second survey participants, a majority ($n = 184$; 86%) have at least one SSO linked to their Google account and 67% ($n = 143$) have at least one third-party app with Google account access. Via the custom browser extension installed by the participants, we observed a total of 1,010 unique SSOs and 455 unique third-party apps accessing participants' Google accounts. For those who have at least one SSO ($n = 184$), the average number of SSOs per participant is 13 (median = 9.5; $sd = 12$). In comparison, those who have at least one third-party app ($n = 143$) have an average number of six third-party apps per participant (median = 3; $sd = 6.7$).

Third-party apps were authorized to access participants' Google accounts for an average of 285 days (median = 142; $sd = 375$). The maximum number of days authorized was 2,519 days and the minimum was one day. The highest number of SSOs linked to a single participant's Google account is 65, and one participant had 49 third-party apps, the most observed. For a list of study participants' most installed apps please refer to Table B.3 in

section B.2.

Associated account access permissions. Among the 1,010 distinct SSOs, we recorded 114 unique associated permissions. Moreover, we cataloged 144 unique permissions requested across the 455 distinct third-party apps. The average number of permissions per SSO was three (median = 2; sd = 1.5); per third-party app, the average number of permissions was three (median = 2; sd = 2.2).

The third-party app with the greatest number of permissions was *Health Sync* with 19 permissions. *Health Sync* was followed by: *autoCrat* (14), *FitToFit* (14), *DocuSign GSuite Add-on* (13), *Zero - Simple Fasting Tracker* (13), and *Yahoo!* (12). Only a few (1%; $n = 12$) SSOs ask for a single permission, while 36% ($n = 166$) of third-party apps have only one permission. Additionally, 78% ($n = 792$) of SSOs have two or fewer permissions while 56% ($n = 255$) of third-party apps have two or fewer permissions.

The most common observed permission was “Create, edit, and delete your Google Play Games activity” ($n = 223$). This was followed by: “See your primary Google Account email address” ($n = 189$), “See your personal info, including any personal info you’ve made publicly available” ($n = 177$), “See, create, and delete its own configuration data in your Google Drive” ($n = 71$), and “Associate you with your personal info on Google” ($n = 44$). See Table B.5 for a list of study participants’ most authorized third-party account access permissions.

4.2.2 Awareness and Understanding

Awareness of Third-party Apps and SSOs. In the second survey, we used the browser extension to show participants their newest, oldest, and a random third-party app. For participants with only two apps, we considered those apps as their oldest and newest. And for participants with only one app, we considered it as their oldest. This results in imbalanced participants with oldest apps $n_{old} = 143$, newest apps $n_{new} = 117$, and random

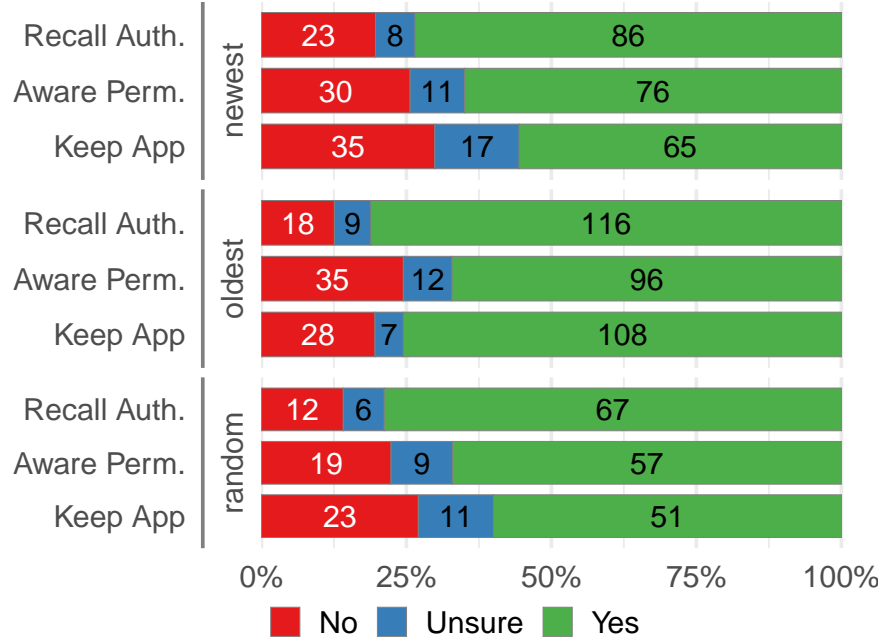


Figure 4.2: Full results of questions (**Q₂₅**), (**Q₂₇**), and (**Q₂₉**). Most participants recall authorizing their apps and are aware that their apps had permissions to access parts of their Google account. Their newest app is the mostly likely to be removed.

apps $n_{rand} = 85$.

Participants were first asked if they recalled authorizing each app (**Q₂₇**). Among participants with at least one app, 33 % ($n = 47$) could not recall authorizing at least one of them. The oldest installed app was recalled the most often 81 % ($n = 116$), followed by the randomly selected app at 79 % ($n = 67$), and finally the newest app at 74 % ($n = 86$). There were no significant differences between the apps shown with respect to positive recall compared to negative or unsure responses ($\chi^2 = 0.27, p = 0.87$). (See Figure 4.3 displays the full details.)

When asked when they last used these apps (**Q₂₈**); over half (51%; $n = 74$) of the participants reported using their oldest app *Today* or *In the previous week*. Whereas 43% ($n = 50$) report using their newest app, and only 34% ($n = 29$) report using their random app, over that same time period. There were no statistical differences (using a Kruskal-Wallace test $H = 2.15, p = 0.34$) between reported last use of apps when comparing newest, oldest, and random apps.

Subsequently, we asked participants if prior to seeing the details about their newest, oldest, and a randomly chosen app in the survey, they were aware that the app had permissions to access their Google account data (**Q29**). Forty-nine percent ($n = 70$) of participants with third-party apps were not aware, for at least one of those apps, that the app had permissions to access parts of their Google account data. Participants were more likely to be unaware of the Google account access permissions of their newest app (35%; $n = 41$), followed by their oldest app (33%; $n = 47$), and random app (33%; $n = 28$). Again, though, there were no significant differences between awareness of the apps ($\chi^2 = 0.032, p = 0.98$). Figure 4.2 shows the full results of app recall and awareness.

In the first survey, when participants were asked if they recalled using Google's SSO (**Q16**), 10% ($n = 42$) responded *No* or *Unsure*. Yet, 16 of the 19 (84%) of those same participants who completed the second survey actually did have a SSO linked with their Google account. We also asked whether they recalled granting a third party access to their Google account (**Q110**), and 47% ($n = 203$) answered *No* or *Unsure*. However, 52 of 95 (55%) of those very same participants who completed the second survey in fact had granted a third-party app access to their account.

Benefits From and Concerns For Third-Party Apps. Participants selected on a 5-point Likert agreement scale to indicate the benefits of their newest, oldest and randomly picked apps (**Q210**). Eighty-two percent ($n = 117$) of participants with apps *Agree* or *Strongly agree* that at least one of their third-party apps is beneficial. The oldest app was reported as the most beneficial with 59% ($n = 84$) who *Agree* or *Strongly agree*, followed by 54% ($n = 63$) for the newest app, and 51% ($n = 43$) for the random app. However, there were no significant differences across responses based on a Kruskal-Wallis test ($H = 2.12, p = 0.34$).

Additionally, participants were asked whether they were concerned about their newest, oldest and random apps having access to their Google accounts (**Q210**). Fifty percent ($n = 71$) of participants with apps *Agree* or *Strongly agree* that they were concerned with at

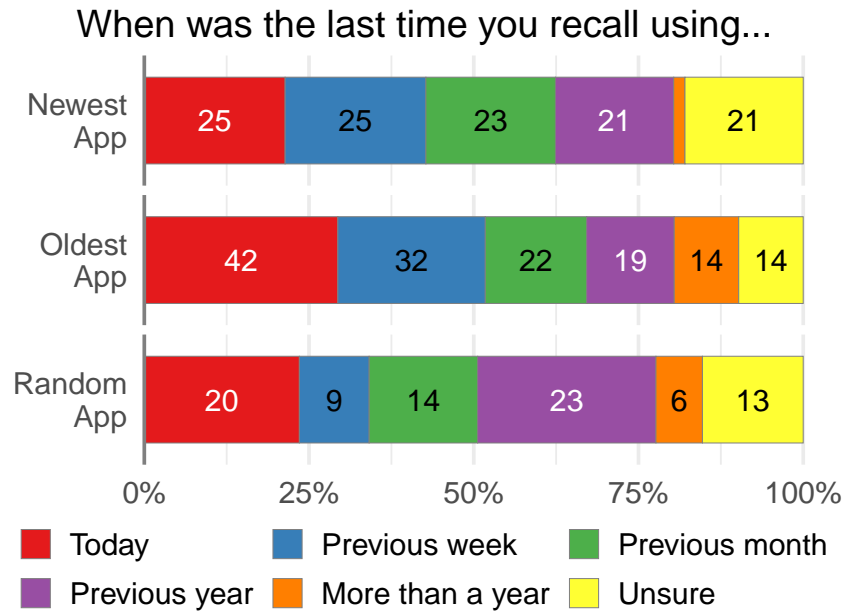


Figure 4.3: Participants’ oldest app was also their most recently used as over half report using their oldest app *Today* or *In the previous week* (Q₂₈).

least one of their third-party apps that can access their Google account. Of the participants who were concerned, 28% ($n = 24$) *Agree* or *Strongly agree* to being concerned with their random app, 22% ($n = 26$) with their newest, and 22% ($n = 32$) with their oldest. The full results for app benefit and concern are shown in Figure 4.4, and again, there were no statistically significant differences ($H = 1.58, p = 0.45$).

Understanding App Access Permissions. We asked participants who have apps to rate their confidence in understanding the permissions held by their third-party apps (Q₂₁₁). Note that each participant had their own set of apps and permissions, so there is an imbalance in the number of participants surveyed for a given permission. Thus, we present permission-specific results as percentages, with full counts in the figures. Thirty-one percent ($n = 45$) of participants had at least one permission that they were *Not confident* that they understood.

For the six most prevalent permissions requested by apps in our study, participants were *Confident* or *Very confident* (over 50 %) in understanding each of them. Participants were the most confident in understanding the permission “See your primary Google Account email

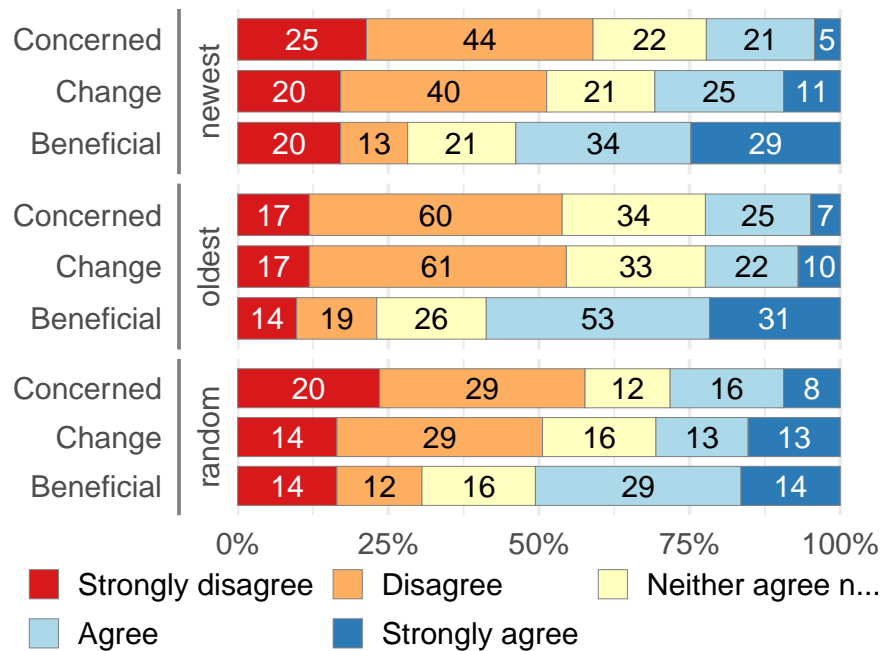


Figure 4.4: Full results of question (Q₂₁₀). Most participants are not concerned about their third-party apps having access to their Google account. Over half of participants do not want to change the parts of their Google account that an app can access. A majority of participants agree that app access to their Google account is beneficial.

address,” with 33 % *Confident* and 39 % *Very confident*. This was also the most common permission surveyed, with 223 occurrences in third-party apps. Conversely, participants were least confident in their understanding of the permission “See your personal info, including any personal info you’ve made publicly available,” with 12 % *Not confident*, 16 % *Slightly confident*. There is also evidence in the qualitative data where a participants note that this permission is confusing because it does not sufficiently detail what information is included in “personal info.” This was the second most common permission surveyed, with 186 occurrences in third-party apps. Results for the top six most prevalent permissions can be found in Figure 4.5. Statistical comparisons were not performed due to the imbalance between groups for which permissions were surveyed.

Necessity of Access Permissions. We asked participants to report the necessity of each permission on a 5-point Likert scale for their newest, oldest and randomly selected third

How confident are you that you understand what each permission allows the app to do?

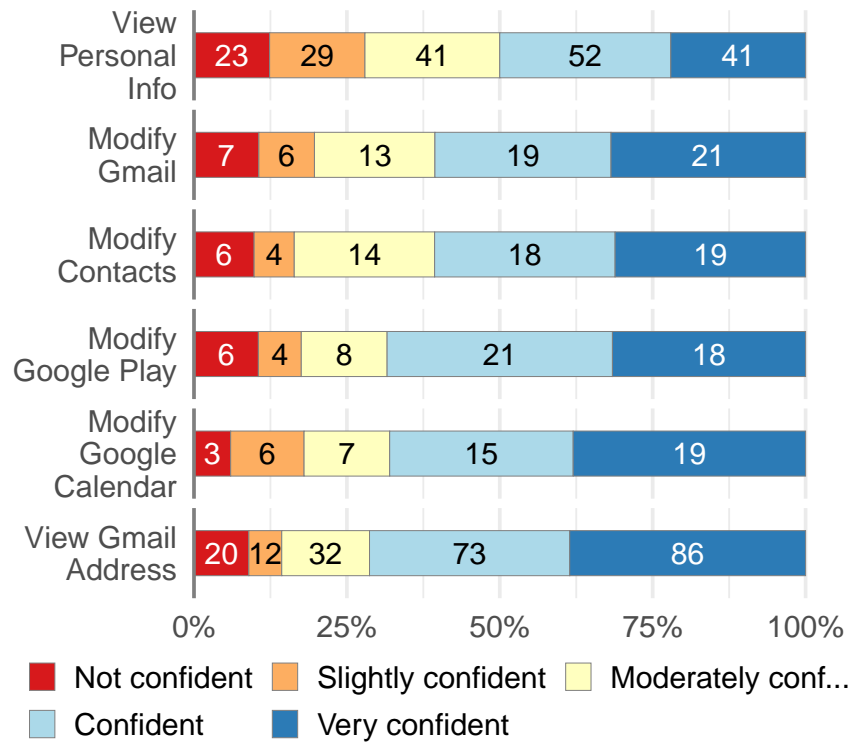


Figure 4.5: Results for (Q₂11) for the top six most prevalent permissions, of which “View Personal Info” was the least understood and “View Gmail Address” was the most.

party app (Q₂12). Sixty-one percent ($n = 87$) of participants had at least one permission that they reported was *Not necessary*.

Among the six most prevalent permissions, participants found the permission “See your personal info, including any personal info you’ve made publicly available” to be the most unnecessary, with 30% who stated that it is *Not necessary* and 31% who said it is only *Slightly necessary*. This is followed by the permission “See, edit, download, and permanently delete your contacts” in which 26% said it was *Not necessary* and 18% said it is only *Slightly necessary*. At least 50% of participants found the remaining prevalent permissions either *Necessary* or *Very necessary*. Participants rated functional permissions, such as “Read, compose, send, and permanently delete all your email from Gmail,” to be more necessary for the app to benefit them than data access permissions, like “See your personal info, including

How necessary do you think each permission is for the app to function in a way that benefits you?

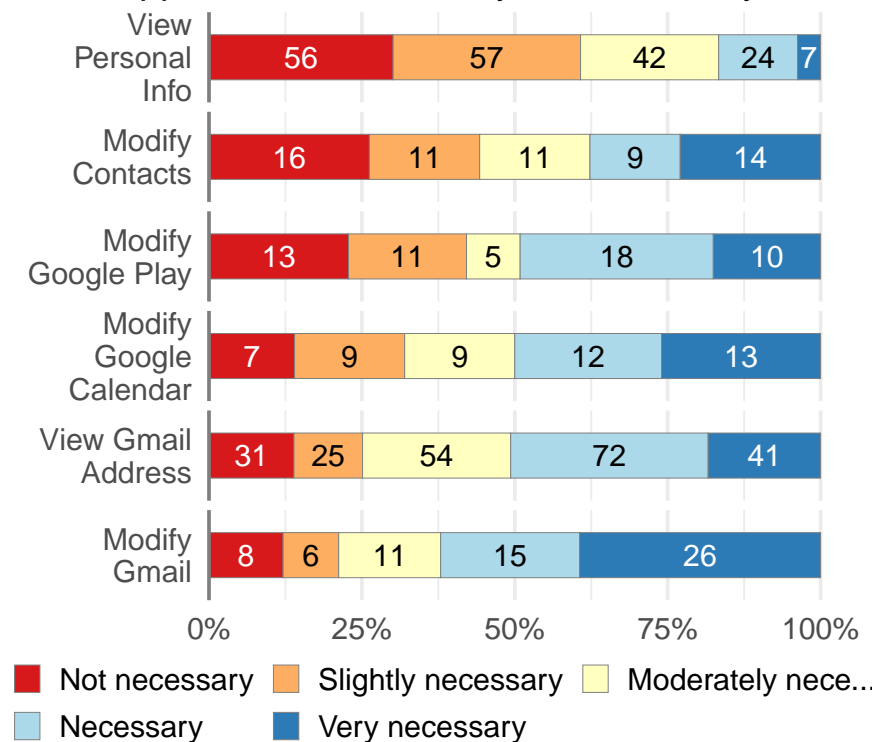


Figure 4.6: Results for (Q₂12) for the top six most prevalent permissions, of which “View Personal Info” was considered the least necessary and “Modify Gmail” was the most.

any personal info you’ve made publicly available.” The top six most prevalent permission results for necessity can be found in Figure 4.6, and again, statistical comparisons were not performed due to the imbalance between groups for which permissions were surveyed.

Concern for Access Permissions. In Q₂13 we asked the level of concern participants have about third-party apps accessing their account through various permissions. Forty-six percent ($n = 66$) of participants had at least one permission that they were either *Concerned* or *Very concerned* about.

For five of the six most prevalent permissions requested by third-party apps observed in our study, over 70% of participants answered *Not concerned* or only *Slightly concerned* about apps on their account having these permissions. The permission “See your personal

How concerned are you about the app accessing your account using these permissions?

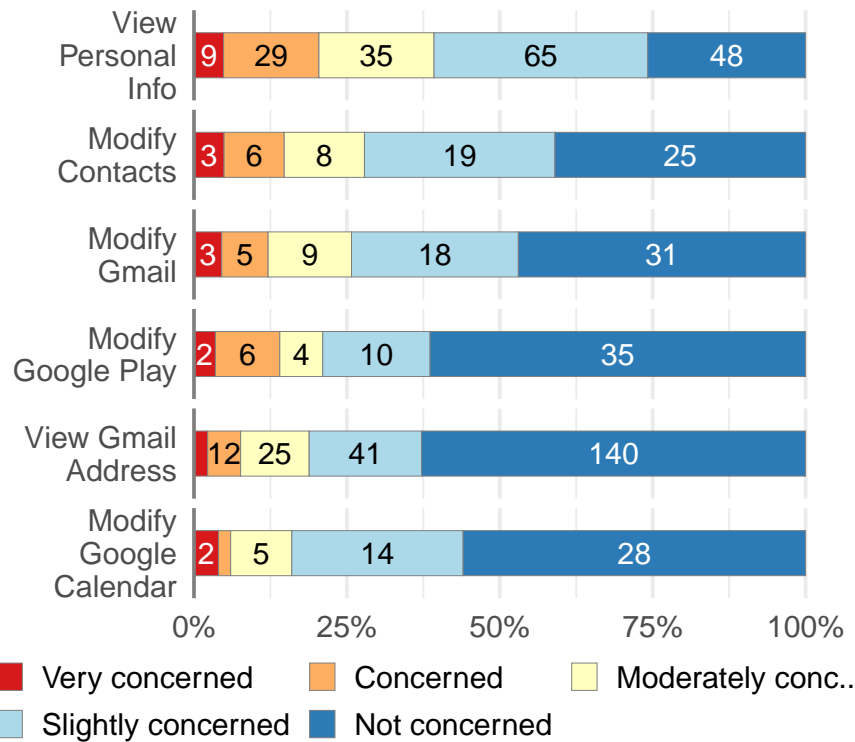


Figure 4.7: Results for (Q₂13) for the top six most prevalent permissions, of which “View Personal Info” was the most concerning and “Modify Google Calendar” was the least.

info, including any personal info you’ve made publicly available” had the highest concern, with 16 % *Concerned*, 5 % *Very concerned*. (See Figure 4.7.)

Reasons for Concern. Participants also provided free responses describing any concerns they have with a third-party app, either the newest, oldest, or randomly chosen, having these access permissions (Q₂14). Qualitative coding of the responses revealed that some participants expressed concern with the permissions held by their apps (newest, $n = 46$; oldest, $n = 90$; random, $n = 33$). The most common reasons for concern were access to personal or sensitive information, unnecessary access, ability to delete, and access to contacts and email. For example, P53 shared, “I don’t want them having access to my personal information” (newest; YouTube on Xbox Live). Examples of concern regarding unnecessary

account access include: “I’m a bit concerned about their access to my Youtube Channel, since it seems a bit unnecessary and excessive” (P170; newest; PlayStation Network) and “Not sure why it needs Google Drive access, it shouldn’t be creating anything in there” (P93; random; Idle Island: Build and Survive). Permissions that include the ability to delete files was often concerning, for instance, “I don’t know if I want them to be able to delete stuff from my Google Drive” (P142; newest; CloudConvert). Access to email and contacts were common concerns; for instance, P63 said, “As with any app that requires having access to send emails, I’m always worried about something unauthorized being sent” (oldest; Boomerang for Gmail). P61 (oldest; Quora) noted:

I didn’t know that they could see and download my contacts. That is a bit concerning because I don’t know what they do with that data.

Participants were also concerned when they could not recall authorizing the access, e.g., “I don’t remember authorizing this app to have access to my Google account” (P10, newest; Email - Edison Mail). Additionally, when participants infrequently used an app but found out it still had access to their account, e.g., “I don’t use it anymore and they still have access to my photos” (P16; random; Chatbooks - Print Family Photos) and “I don’t use the Google nest hub anymore, so it shouldn’t have access to my full account” (P12; oldest; Google Nest Hub). A number of participants complained that they had deleted the app from their device but it still had access to their accounts, as when P137 shared, “I have removed the app from my phone and I don’t see why the app still has to have these permissions” (newest; Adidas Training), and P26 noted, “I was unaware these permissions were still on the app as I’ve deleted the app” (newest; Linkt).

Many participants stated that they were unconcerned by the app permissions (newest, $n = 62$; oldest, $n = 44$; random, $n = 41$). The most common reasons for the lack of concern was the necessity of the permissions, and the limited access that the permissions provided. For instance, P131 said “I don’t think it has too many permissions and nothing seems unreasonable, so I’m okay with this” (random; Lumin PDF), and P158 stated “It is a game

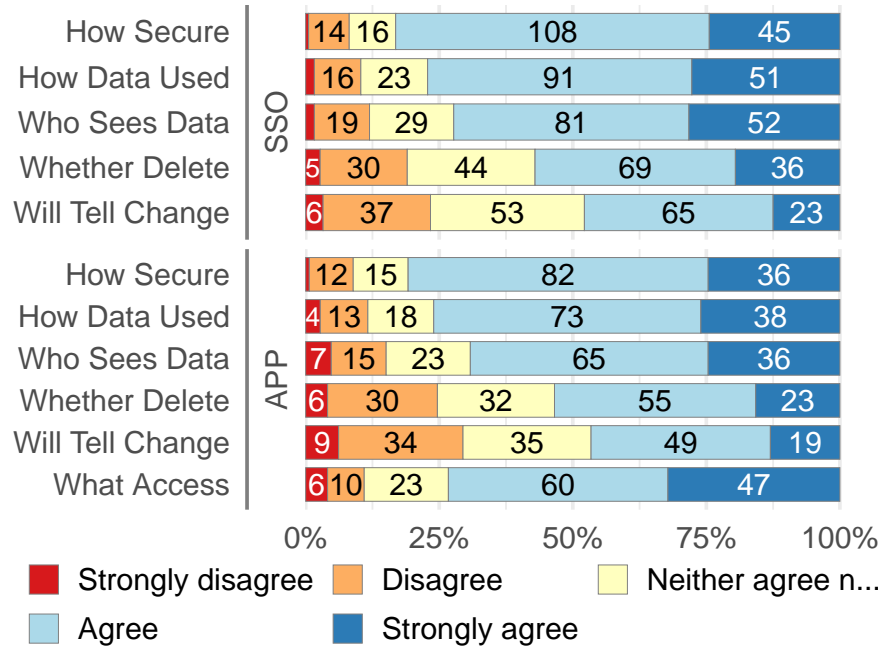


Figure 4.8: Full results of questions (Q₂1) and (Q₂3) in which participants are asked what they consider before granting Google account access to a third-party APP or SSO.

and both of the permissions are valid” (random; Starlost - Space Shooter). Some were unconcerned because they trust the app, e.g., “I trust the service to keep my information secure” (P148; newest; Amazing Marvin). For others, the trust originates with the company who makes the app, e.g., “I’ve never worried about Apple invading my privacy” (P189; random; macOS).

4.2.3 Granting and Reviewing Account Access

Considerations When Granting App/SSO Access. Google’s documentation [60] advises that users consider the following five factors before granting a third-party access to their Google account: (i) security, (ii) data use, (iii) data deletion, (iv) policy changes, and (v) data visibility.

On a five-point agreement Likert-scale, we asked the participants with apps if they considered these factors before granting a third-party app access to their Google account (Q₂3) and before using their Google account to sign into a service via SSO (Q₂1); those

results are presented in Figure 4.8.

The most considered factor for third-party apps was the security of the app or website (security), in which 83 % ($n = 118$) *Agree* or *Strongly Agree*. The next factor was how the app or website will use your data (data use) where 78 % ($n = 111$) *Agree* or *Strongly Agree*. This was followed by who else can see your data on the app or website (data visibility), in which 71 % ($n = 101$) *Agree* or *Strongly Agree*. Next was whether you can delete your data from the app or website (data deletion), where 55 % ($n = 78$) either *Agree* or *Strongly Agree*. The least considered factor was whether the app or website will tell you if something changes (policy changes), where fewer than half ($n = 68$; 48 %) *Agree* or *Strongly Agree*.

Among participants with SSOs ($n = 184$), the data shows they considered similar factors as those for third-party app account access. Again, the most considered factor was how secure is the app or website (security), in which 83 % ($n = 153$) *Agree* or *Strongly Agree*. The next most considered factor was how the app or website will use your data (data use) where 77 % ($n = 142$) *Agree* or *Strongly Agree*. This was followed by who else can see your data on the app or website (data visibility), in which 72 % ($n = 133$) *Agree* or *Strongly Agree*. Next was whether you can delete your data from the app or website (data deletion), where 57 % ($n = 105$) either *Agree* or *Strongly Agree*. The least considered factor was whether the app or website will tell you if something changes (policy changes), where fewer than half ($n = 88$; 48 %) *Agree* or *Strongly Agree*.

We used a Mann-Whitney U-test to compare each of the considerations, comparing SSO to third-party access (see top of Figure 4.8). We did not find any significant differences, suggesting that participants view SSOs and third-party apps accessing their Google accounts in similar ways when determining if they should grant that access. More detail on SSO considerations is in the next section.

In open-response questions, we asked participants to provide more details on what they consider before granting a third-party app access to their Google account or use SSOs (Q₁₂, Q₁₃). Participants often (Q₁₂, $n = 42$; Q₁₃, $n = 43$) responded that they

consider what permissions the third-party would obtain, e.g., “What capabilities I was giving the third-party” (P362). Security (Q12, $n = 33$; Q13, $n = 23$) and privacy (Q12, $n = 22$; Q13, $n = 28$) were common considerations. For instance P160 noted, “Whether it was secure and could I trust it,” and P283 added, “I’m always worried about my privacy anytime a app [sic] asks me for that information.” Still, many participants (Q12, $n = 31$; Q13, $n = 10$) had no considerations. For instance, P24 shared, “It was a pop up so I didn’t consider it much at all.”

One particularly interesting theme that emerged from open coding was the transfer of trust from the Google brand name to third-party apps and SSOs when participants considered granting access (Q12, $n = 7$; Q13, $n = 4$). That is, participants were more likely to trust a third-party service because it was using Google, and they trusted Google, irrespective of the nature of the third-party service. For example, P278 declared, “That it was okay since Google was allowing it,” and P297 added, “It must be okay since it partnered with Google.” P117 provided another example:

I would consider nothing again, I probably put too much trust in Google and it’s become a crutch at this point, I would easily allow it to be used in 3rd party situations.

and P161:

Nothing! Like before, I generally trust anything that leads to that Google SSO page.

When considering granting SSO access (**Q18**), many participants ($n = 55$) considered ease of use and convenience before signing in, and a common theme was the ease of reusing their Google account login credentials versus creating a new account on a third-party app or website, e.g., “It is easier and more convenient than making a brand new account to some third party website” (P21). There were also many ($n = 42$) who were unconcerned and had few considerations, such as P99 who said, “I didn’t consider much, I use Google sign in pretty regularly,” and P117 who stated, “I didn’t consider much, my Google account has always been good to me and offered ease of use with many things.”

Participants ($n = 27$) also shared concerns such as what information would be accessible to the third-party app or service; for instance, P90 responded, “If that service has access to information associated with my Google Account,” and P91 replied, “I considered what that app would have access to if I signed in through there.” Security ($n = 29$) and privacy ($n = 13$) were common considerations. For example P80 shared, “I was worried about security of my Google Account,” and P271 added, “The private data that this application was going to read, in other words, I worry about my privacy.”

Some participants ($n = 12$) described a trade-off between information sharing and convenience. For instance, P16 shared, “The effort of making a new account versus sharing my google info.” Trust of the website or service being signed into is also a common theme ($n = 26$), e.g., “I consider if the website is trustworthy and if I can trust signing in using my Gmail account in their website” (P68).

Reasons For Authorizing Account Access. We asked participants what was the purpose of allowing third-party access (**Q11**). The purpose for many ($n = 45$) participants was the utility that the app provided, such as email and contact management, file transfer, and synchronizing data between devices. For instance, P37 explained, “I gave a ringtone app access to my contacts and messages so that it can change the notification/ringtone sounds” and P210 shared that the app’s purpose was “Allowing me access to multiple e-mails from the same app on my Windows computer.” Another popular ($n = 38$) purpose was calendar management, e.g., “Zoom, to allow it to add meetings to my calendar” P25. Gaming was a common ($n = 38$) purpose. For example, P355 responded:

I allowed access to my Google account for a lot of games, because I can get achievements to be displayed on my account with Google, and also it is usually an easy way to recover or save data between devices.

Some participants ($n = 15$) just could not recall the purpose, like P404 who shared, “I don’t really remember but I do remember encountering this type of screen in the past.”

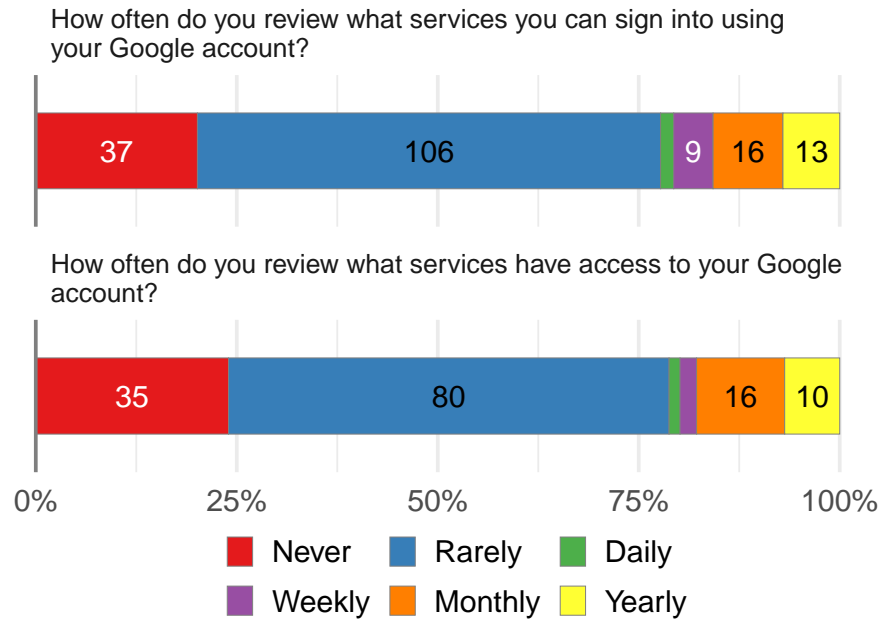


Figure 4.9: Over 75 % of participants stated that they *Never* or *Rarely* review what services they can sign into using (Q₂₂) or have access to their Google account (Q₂₄).

User Review of Apps With Account Access. We find that participants rarely or never review the services they can sign into using their Google account or the apps that have access to their Google account. We asked participants how often they review the services they can sign into using their Google account (Q₂₂). A majority ($n = 106$; 58 %) *Rarely* review their services, 20 % ($n = 37$) *Never* review, and 16 % ($n = 29$) review them *Monthly* or *Yearly*. Among participants with apps, we also asked how often they review the services that have access to their Google account (Q₂₄). Again, a majority ($n = 80$; 56 %) *Rarely* review their services, 24 % ($n = 35$) *Never* review, and 18 % ($n = 26$) review them *Monthly* or *Yearly*. Figure 4.9 shows the full results of these questions.

For each of their newest, oldest and random app, we also asked participants if they would like to change which parts of their Google account that the third-party app can access (Q₂₁₀). 52 % ($n = 74$) of participants *Agree* or *Strongly agree* they want to change which parts of their Google account are accessible for at least one of their apps. The most agreement for changing access was for the random app, in which 30% indicated they *Agree* (15%; $n = 13$)

Table 4.2: Binomial logistic model to describe which factors influenced the preference to remove an app (*Remove* responses to question **Q₂₅**). The Aldrich-Nelson pseudo $R^2 = 0.52$.

Factor	Est.	OR	Pr(> z)	
(Intercept)	-1.02	0.36	0.134	
Participant's newest app	0.22	1.24	0.687	
Participant's oldest app	-0.07	0.93	0.904	
Recall = <i>Yes</i>	0.92	2.50	0.179	
Aware app permissions = <i>Yes</i>	-1.15	0.32	0.058	.
Last use $\in \{day, week, month\}$	-1.76	0.17	<0.001	***
Access benefit $\in \{Agree, Strongly Agree\}$	-1.80	0.17	0.001	**
Access concern $\in \{Agree, Strongly Agree\}$	0.43	1.53	0.443	
Access change $\in \{Agree, Strongly Agree\}$	1.75	5.76	0.001	**
Number of permissions > median	0.05	1.05	0.909	
Time since install < 3 months	0.84	2.32	0.123	
Time since install > 2 years	0.26	1.30	0.689	
Signif. codes: *** $\hat{=}$ < 0.001; ** $\hat{=}$ < 0.01; * $\hat{=}$ < 0.05; . $\hat{=}$ < 0.1				

or even *Strongly agree* (15%; $n = 13$). This was followed by the newest app, where 21% ($n = 25$) *Agree* and 9% ($n = 11$) *Strongly agree*. The oldest app had the least agreement for a change in access with only 15% ($n = 22$) *Agree* and 7% ($n = 10$) *Strongly agree*.

Keeping or Removing Apps. For each of the specific apps shown—newest, oldest, and randomly chosen—we asked participant if they would like to keep or remove the app or are unsure about what to do (**Q₂₅**). 43 % ($n = 62$) of participants with third-party apps wanted to remove at least one of those apps. Due to private data aggregation during the survey, we only linked the keep/remove preferences for the newest, oldest, and random app that were specifically reviewed by the participants. For their newest app, 56 % ($n = 65$) of the participants said they want to keep it, 30 % ($n = 35$) chose to remove, and 15 % ($n = 17$) answered unsure. Many more participants (76 %; $n = 108$) responded to keep their oldest app. While 20 % ($n = 28$) wanted to remove, and 5 % ($n = 7$) were unsure. 60 % ($n = 51$) of participants wanted to keep their randomly selected app, 27% ($n = 23$) answered to remove, and 13% ($n = 11$) were unsure. However, a Kruskal-Wallace test found no significant differences between the three apps shown ($H = 0.49, p = 0.77$). Full results in Figure 4.2.

We performed logistic regression to determine factors that would lead a participant to remove a third-party app access from their Google account (see Table 4.2). We controlled for repeated measures by adding random intercepts to the model, as each participant provided up to three apps they wished to keep or remove. We found a significant correlation with participants who want to change which parts of their Google account the app can access ($\beta = 1.75, OR = 5.76, p = 0.001$), where those participants were $5.8\times$ more likely to want to remove the app access. We found a significant correlation with participants who have used the app within the last month ($\beta = -1.76, OR = 0.17, p = < 0.001$), and those participants are $5.9\times$ more likely to want to keep the app access. We also found significant correlation with participants who found the app beneficial ($\beta = -1.80, OR = 0.17, p = 0.001$), where they too are $5.9\times$ more likely to want to keep the app access. These findings suggest the importance of app usage frequency in account access authorization.

4.2.4 Reflection and Features

Understanding Account Linking and Access. We directed all participants in the second survey ($n = 214$) to explore their “Apps with access to your account” page. We then asked whether the page helps them to better understand which third-party apps and websites are linked to their Google account (**Q₂15**); nearly all participants ($n = 204$; 95 %) agreed that it help. And when asked if the page helps to better understand which parts of their Google account third-party apps can access (**Q₂16**), again nearly all participants ($n = 198$; 93 %) agreed. This suggests that the management page has the potential for being an important tool.

Change Settings and Review Apps. We asked participants to indicate whether they intend to change any settings after seeing their “Apps with access to your account” page (**Q₂17**), and roughly half ($n = 105$) affirmed that they would change settings. Over 70 % ($n = 152$) indicated they plan to review third-party apps in six months (**Q₂18**). Refer to Figure 4.10 to

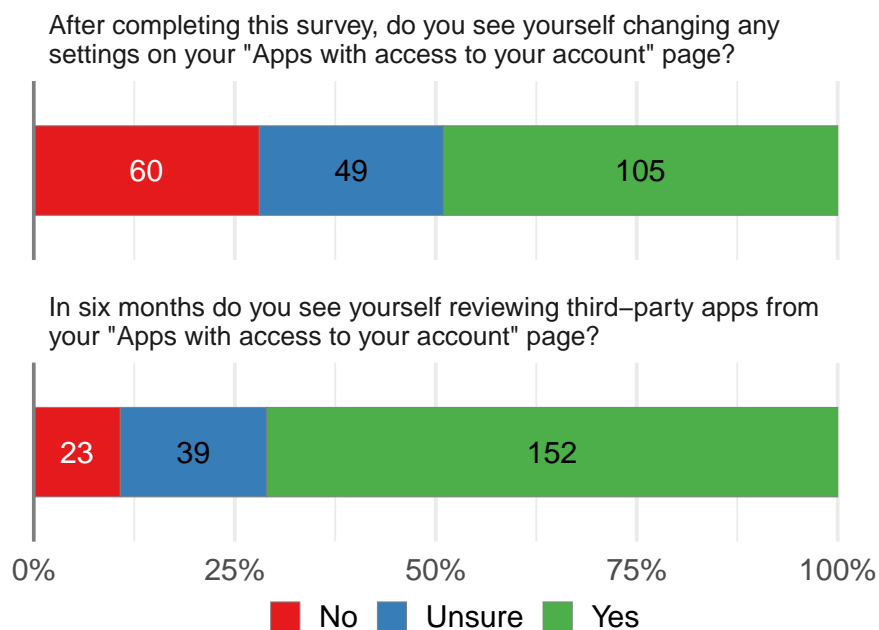


Figure 4.10: Roughly 50 % of participants stated that they would change their settings after the survey (Q₂17), and a majority would review third-party apps in six months (Q₂18).

see the full results of these questions.

If a participant affirmed they would change a setting, we asked them to tell us which settings they would change (Q₂19). Many participants ($n = 76$) wanted to remove access from one or more apps. They often ($n = 41$) mentioned unused apps; for instance, P23 said:

I would definitely remove many of the apps that I do not use anymore. They absolutely do not need to be linked to my Google account anymore.

Participants also wished to remove apps they no longer recalled authorizing, e.g., “There are apps I do not use and do not recall allowing access to my Google account” (P24).

Some participants ($n = 27$) wanted to change specific permissions access, something not allowable with the current interface. The most common permissions mentioned included contacts, account info, delete or modify files, and “unnecessary permissions.” For example, P189 said, “I would remove Dropbox’s access to my contacts,” and P192 noted, “... and maybe restrict Streak’s access to my Google Drive.” Participants ($n = 8$) mentioned protecting their personal information as a reason for the settings change, e.g., “I would limit the

amount of information different apps can view - especially with my personal information” (P145).

If a participant affirmed they planned to review third-party apps in six months, we asked them to tell us what they would look for (**Q₂20**). Many participants ($n = 43$) would look for apps that they no longer use, such as P137, who replied “I would look for apps that I no longer use and if they still have access to my account and to what parts of my account.” Some participants ($n = 22$) said they would look for new apps that have access to their account, e.g., “If any new apps are listed that I don’t remember approving” (P169). Others ($n = 22$) wanted to review changes to the list of apps with access to their Google account. For instance, P138 explained:

I would look for any unexpected or new third parties and just make sure that all of them have very limited access to my personal information. I’d maybe continue to review this page over time to make sure nothing has changed.

Some wanted to review changes to the existing apps, like P180 who shared, “Check to see if anything has changed like they got more access without saying anything.” Participants ($n = 8$) also wanted to review how much access was allowed to third-party apps. For example, P210 stated:

I would want to know exactly what those apps have access to. Actually wish there was more information there. Seems like it is a little generic.

New Features and Design Changes. In an open-ended response question, we asked what new features participants would like to add to the “Apps with access to your account” page (**Q₂21**). Many participants ($n = 61$) wanted the page to display more detailed information and for the account access to be more transparent, e.g., “An option to have even more detail of what was exactly being accessed” (P68). A common ($n = 19$) request was an app usage or data access log, like when P17 replied:

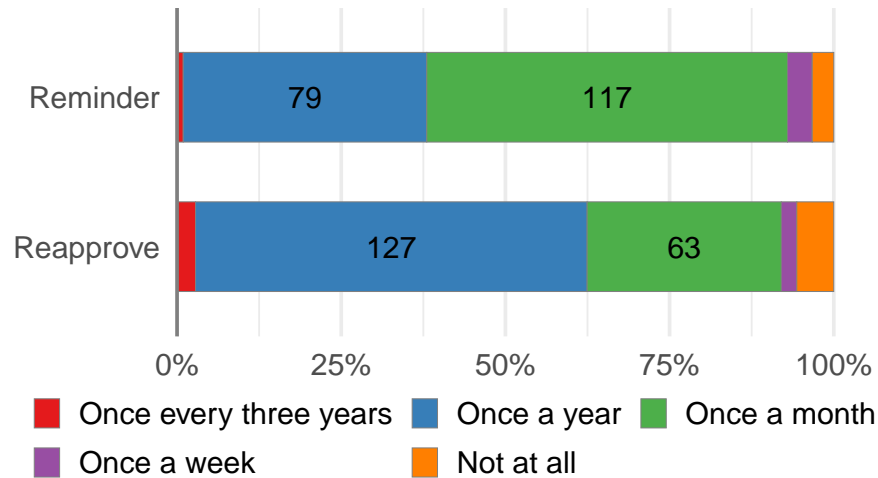


Figure 4.11: A majority of participants stated that they wanted an email reminder to review their “Apps with access to your account” page *Once a month* (Q₂₂) and to reapprove apps *Once a year* (Q₂₃).

I think it would be useful for them to show when I last used an app and when the app last used my data, to see if the app is using my data even when I have not used the app in a while.

Another request ($n = 12$) was detailed permission explanations, for instance when P174 demanded,

DETAILED information about what is available to third party apps. Not just “access to your personal information,” but “access to: your full name, age, date of birth, street address.”

Permissions level control is also desired, e.g., “The ability to remove certain accesses that I don’t think the app needs” (P143), and some requested access timeouts, e.g., “Auto removal of apps after 3 months of no use” (P156).

We asked participants how often they would like to be reminded if Google were to provide an email reminder to review their “Apps with access to your account” page (Q₂₂). Most participants ($n = 117$; 55 %) responded *Once a month*, another 37 % ($n = 79$) want a reminder *Once a year*.

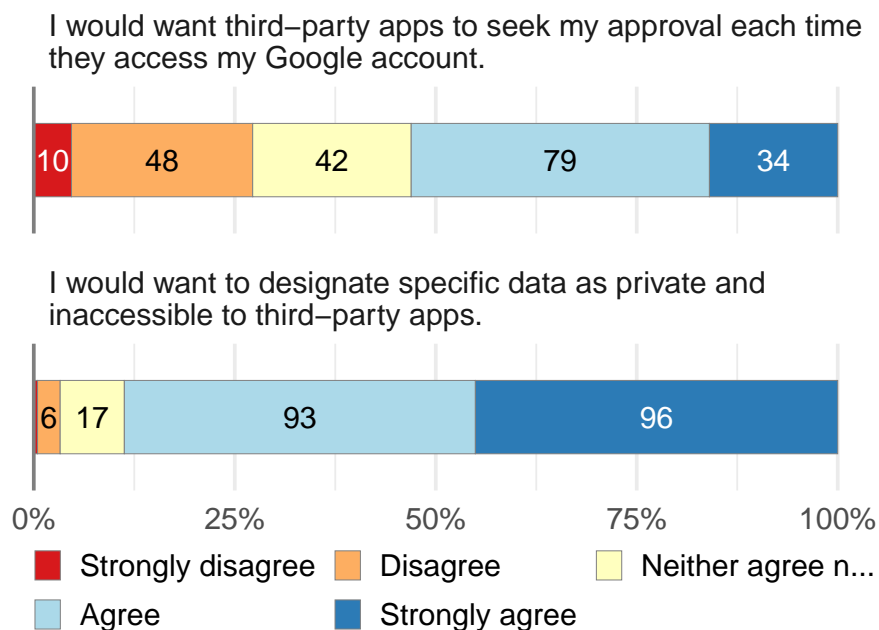


Figure 4.12: Over half of participants *Strongly Agree* or *Agree* they want third-party apps to seek approval each time they access their Google account (Q₂₂₅), and roughly 90 % of participants *Strongly Agree* or *Agree* they want to designate specific data as private and inaccessible to third-party apps (Q₂₂₆). One participant managed to submit a non-response, and is excluded from this data.

Reapproving Apps. Additionally, we asked participants if Google required them to reapprove the third-party apps with access to their account, how often would they want to provide reapproval (Q₂₂₂). A majority of participants ($n = 127$; 59 %) want to reapprove apps *Once a year*, while 29 % ($n = 63$) want a reminder *Once a month*. Please see Figure 4.11 for the full results of these questions.

When we asked participants if they would want third-party apps to seek approval each time they access their Google account (Q₂₂₅) over half ($n = 113$; 53 %) agreed they want third-party apps to seek approval each time they access their Google account, while 27 % ($n = 58$) disagreed. We also asked participants if they want to designate specific data (e.g. certain emails, individual contacts, particular calendar events) as private and inaccessible to third-party apps (Q₂₂₆), and most participants ($n = 189$; 89 %) agreed, few ($n = 3$; 1 %) disagreed. Please refer to Figure 4.12 for the full results.

4.3 Discussion and Conclusion

In this section we offer conclusions and recommendations for managing third-party apps and SSOs that access users' Google accounts. We first discuss issues with handling disused third-party access and the potential to have cascading removal. Next, we focus on issues with transference of trust from Google to third-party apps, and finally, we offer some design improvement for “Apps with access to your account” based on our observations and qualitative feedback.

Handling Stale Account Access. This research identifies the need for limiting account access for unused, forgotten, or removed third-party apps. Participants either had not recently used or were unsure when they had last used one in five third-party apps in our study. Additionally, most participants (80%) rarely or never reviewed the third-party account access. At the same time, nearly half of participants wanted to modify access after the study and over 70% wanted to do so within six-months. Most participants indicated they would remove unused apps or apps they no longer recall authorizing, but the current interface for reviewing third-party apps does not provide a record of last use.

Participants also expressed strong support (95%) for email reminders to review their third-party apps at least once a year. Moreover, participants highly favored (91%) a requirement to reapprove account access at least once a year. Google could send email reminders to all users to review “Apps with access to your account” with additional details provided regarding unused third-party apps. Review request could also be more directed, perhaps asking users to review a single app at a time so as not to overburden. Another possibility is to auto-expire apps that have not been recently used or did not recently make an API call. Facebook implemented a similar policy following the Cambridge Analytica scandal [77], to expire app access periodically and then require users to reauthorize.

Cascading Removal. Participants were surprised to find apps in their “Apps with access to your account” page that they believed were removed in other places. For example, many smartphone applications leverage a Google Account for functionality and authentication, and in these case, participants who removed that smartphone app believed they also remove the third-party app access.

While expiring third-party apps, see above, would help address this issue, cascading removal could provide additional support for users who attempt to manage access to their Google account. A third-party app, or the user’s account, can send a notification when a secondary application that leverages the third-party app is removed from a user device, which in turn allows the user to deactivate the third-party app access.

All Or Nothing Permissions. Users must approve *all* permission requests for third-party apps or none, and many of permission requests may not be fully understandable, contextualized, or presented at the time of the access request. After granting access, the “Apps with access to your account” offers no permission-level control to limit account access; the only option is to completely remove access.

Even without fine-grain controls, most participants in our study find most permissions to be necessary for the functionality of their apps; however, they consider some permissions unnecessary and concerning, especially those that allow access to view personal information that may include data beyond just name and email address. Our qualitative data shows that this practice is forcing users to consider a tradeoff between their personal privacy and the benefits of third-party apps or the convenience of SSOs.

Furthermore, when prompted to describe new features, participants stated a desire for permission level controls, and we recognize such a feature could have negative impacts on the functionality of third-party apps. However, we recommend that Google provide permissions level control for users’ personal data, for instance: “See your gender,” “See your age group,” “View your street addresses,” “See and download your personal phone

numbers,” “See your personal info,” and “Associate you with your personal info on Google,” which likely are not used for app functionality.

Trust Transference. Google is a highly ranked and trusted brand [111], so trust in Google transfers to Google products and services, including third parties’ apps that access Google accounts and services. We observed this in many open-ended responses from participants that assumed (incorrectly) an implied partnership between Google and third-party apps.

To Google’s credit, it attempts to mitigate this effect by noting that users should “Make sure you trust [*third-party app name*],” on the access authorization prompt. However, this does not appear to be a sufficient intervention, and more so, it places a potential undue burden on users to be able to properly differentiate. This further suggests that better messaging, in the form of nudges or reminders to review third-party apps, or other automated mechanisms should be put in place to assist users to better manage apps and services that access their Google account.

Improving App Transparency Tool. We also note that there are number of possible design improvements to the existing “Apps with access to your account” page based on participant recommendations offered during the study and our findings. Nearly 90% of participants suggested that Google should provide users with the ability to designate specific account data as private and inaccessible to third-party apps. For instance: certain emails, individual contacts, and particular calendar events. Moreover, adding a recent activity log that includes data access details would allow users to determine the frequency with which they use a third-party app and to determine if it is accessing data while appearing inactive.

Improvements to the permission descriptive text could also include specific details about which parts of the users’ Google account data are accessible. For example, instead of “See your personal info,” the personal info available to see should be enumerated, e.g., full name, email address, age, street address, and profile picture. Contextual information could also be included, such as identifying the reasons a third-party app is requesting a certain permission.

For instance, “Zoom has access to your Google calendar to schedule and modify Zoom meeting times,” instead of simply noting it has access to the calendar.

Chapter 5: Longitudinal Analysis of Privacy Labels in the Apple App Store

The ubiquitous surveillance and data collection regime embedded within modern web and mobile ecosystems has led to several interventions meant to empower users to make choices to restrict and manage their privacy. Most prominent are privacy policies as free-text explanations of what data services collect and how that data is used. Privacy policies are often complicated and difficult for users to understand and use for decision making [30, 98]. Privacy nutrition labels [83] offer an alternative approach that is modeled after food nutrition labeling [54]. Like a food label, a privacy nutrition label describes the data collection and usage practices of a service. Privacy nutrition labels (or *privacy labels*) have been proposed for a range of products, most notably for the Internet of Things (IoT) [41, 83, 85], where interfaces and interactions can be limited.

In December of 2020, Apple made privacy labels mandatory for all new and version-updated apps on the Apple iOS App Store. The structure of Apple’s privacy labels (see Figure 5.1 and Figure 5.2) standardizes information that was previously difficult to obtain for users, such as the type of data collected by an app (e.g., *Email Address*, *Payment Info*, *Precise Location*), the purpose of the collection (e.g., *App Functionality*, *Product Personalization*), and the ways data will be used (e.g., tracking users across apps and websites, linking to users’ identities). In April of 2022, Google followed and announced its own form of privacy labels for the Play Store [63].

The standardization of privacy labels offers a unique opportunity for a comprehensive study of the self-reported data collection and use policies of the entire iOS app ecosystem. We conducted a 36-week (from July 15 2022 to March 16, 2022) longitudinal analysis of all 1.6 million apps in the iOS App Store, collecting their privacy labels and metadata while performing a comprehensive analysis of how apps are coming into compliance with this new policy.

At the end of the measurement period, in March 2022, nearly two years after Apple announced the new policy, only 60.5% ($n = 952,922$) of apps in the app store have a privacy label, but we observed a slow and steady increase of 0.5% ($n = 8,068$) of apps per week coming into compliance. From the start of collection, there was an 18.4% increase in apps with labels (670,547 vs. 952,922). However, this increase was primarily driven by new apps added to the store that are required to have a label, rather than older, legacy apps that are being updated or voluntarily adding labels.

The most frequently reported purposes for data collection in the privacy labels are *App Functionality* and *Analytics*. *Third Party Advertising* was only noted 18% and 20% of the time when an app had *Data Linked to You* or *Data Not Linked to You*, respectively, despite being a very common reason for data collection. *Identifiers* and *Usage Data* were the most common data categories used to track users, while *Contact Info* and *Identifiers* were most commonly linked to users. *Diagnostic* and *Usage Data* are the largest share of data categories of data collected but *not* linked to users.

We also compared apps based on a number of criteria, including their (age-based) content rating. Fifteen percent of apps with a content rating of 4+ have a privacy label indicating *Data Used to Track You*, as do fifty percent of apps with a content rating of 9+. Such apps may be directed at children and would be subject to data collection and tracking standards in line with Children's Online Privacy Protection Act (COPPA), requiring parental consent for data collection and tracking.

In comparing paid vs. free apps, more free apps use privacy labels that indicate data collection and tracking. Forty-two-percent of free apps with in-app purchases have a *Data Used to Track You* label, 41% a *Data Linked to You* label, and 61% have a *Data Not Linked to You* label, while only 24% report *Data Not Collected*. Of paid apps, 2.4%, 4.2% and 14% for the same labels, respectively, and 84% indicate *Data Not Collected*. This data likely reflects additional revenue streams from free apps in targeted advertising and/or selling user data.

Importantly, privacy labels are self-reported and not validated by Apple. An app’s label is entirely at the discretion of the developers, and we observed different labeling depending whether the developers were forced to add labels, like a new or updated app, as compared to voluntarily adding labels. Fifty percent of developers forced to add labels opted for *Data Not Collected* but only forty-one percent did so when voluntarily adding labels. Many developers may see privacy labels as an obstacle to the ultimate goal of adding an app to the store, rather than a mechanism to communicate about privacy practices.

We also observed very few shifts in privacy labels. Only 13,785 apps (of the nearly 1M apps with labels) had a label that changed during the 36-week observation. The first label applied will likely persist, and when privacy labels for apps do change, they tend to get more invasive, including additional data categories collected/tracked. As labels are not validated by Apple, it is difficult to know the veracity of labels, but as changing a label is strictly a voluntary act on the part of the developer, it is likely that many apps are significantly under-labeling.

Through this longitudinal measurement, we expose the scope and prevalence of data tracking in the iOS ecosystem, and it likely offers a skewed view of privacy practices. The truth may be far worse, as privacy labels lack any form of validation. We found a number of apps that obviously must collect user data, but failed to report so in their privacy labels. While future work in app analysis is needed to validate privacy labels, we also argue that comprehensive and continual transparent measurements at scale can provide important context and potentially some accountability for developers for the privacy labels used. This is particularly pressing as privacy labels are becoming the de facto standards across the mobile app marketplaces beyond Apple.

5.1 Background

The new Apple privacy label is similar in style and content to the “Privacy Facts” label developed by Kelly et al [85]. The structure of Apple’s privacy labels is hierarchical (see

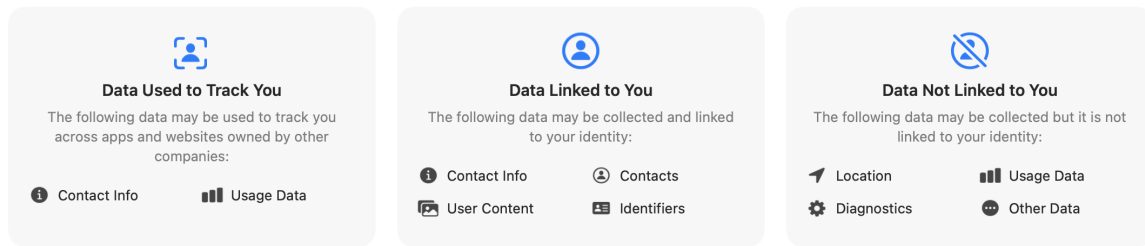


Figure 5.1: An illustrative example of a privacy label from the Apple App Store.

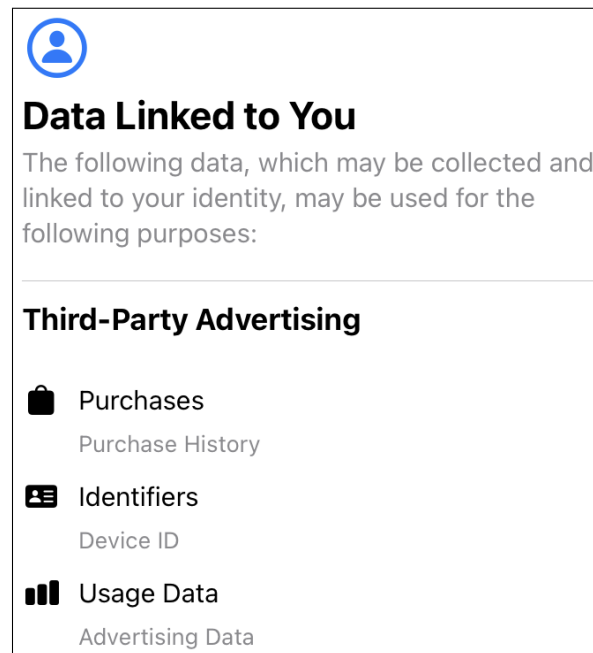


Figure 5.2: An illustrative example of the privacy label details from the Apple App Store. The details display the *Purposes* for the data collection and the detailed information about the *Data Types* collected.

Figure 5.3 for details) and are the combination of four sections of information. In the top level (to the left in the diagram) of the label hierarchy are four distinct *Privacy Types*, three of which describe ways of using data. An app's privacy label may contain a combination of one, two, or all three of these types. The fourth *Privacy Type*, entitled *Data Not Collected*, is displayed with an image of a blue checkbox and indicates that the developer does not collect any data from this app. *Data Not Collected* is mutually exclusive with the other three *Privacy Types* (see Figure 5.5), and an app labeled with *Data Not Collected* cannot list other *Privacy Types*.

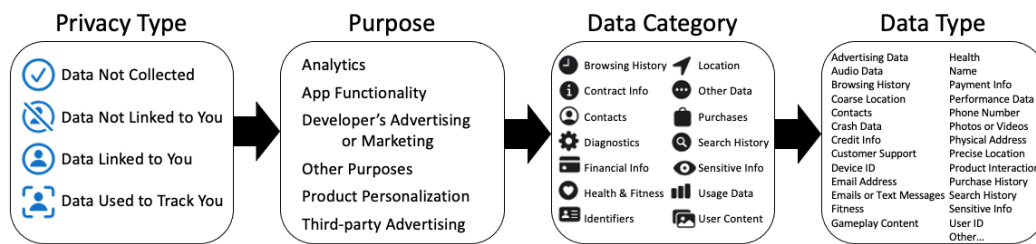


Figure 5.3: A privacy label consists of four hierarchical levels of information. The *Privacy Type* broadly identifies how the data collected will be used. The *Purpose* provides more detail on how each data type is used. The *Data Category* is a categorization the *Data Type* which is a detailed description of the type of data collected.

Among the three *Privacy Types* signifying data collection, the first, entitled *Data Used to Track You*, indicates that data collected may be used to track users across apps and websites owned by other companies, including sharing data with third-party advertising networks and data brokers. The second privacy type, entitled *Data Linked to You*, indicates that data may be collected and linked to the user's identity. The third, entitled *Data Not Linked to You*, indicates that data may be collected but it is not linked to the user's identity.

The next level down in the label hierarchy (moving right in the diagram) are the *Purposes* which group data into six purposes for which data is collected, such as *Third-Party Advertising*, *Product Personalization*, or *App Functionality*. *Purposes*, which are only provided for the *Data Linked to You* and *Data Not Linked to You Privacy Types*. The purpose for *Data Used to Track You* data collection is to link with third-party data for advertising or advertising measurement purposes, or to share data with a data broker.

The *Data Categories* are the next level, describing 14 categories of data collection, such as *Contact Info*, *Location*, and *Purchases*. And finally, at the bottom level (to the right in the diagram) are the *Data Types*, which provide the most detailed grouping of the data collected into 32 descriptive types, such as *Coarse Location*, *Precise Location*, *Gameplay Content*, or *Emails or Text Messages*.

In Figure 5.1, we provide an example of a privacy label as displayed to users. Each box reports the *Privacy Type*, with a short description of that type, and then within the box, each

of the *Data Categories* is presented. When one clicks on a given category, a popup screen is displayed (see Figure 5.2) with details about the *Purposes* and specific *Data Types*.

5.2 Methodology

We started our weekly data collection on July 15, 2022. Each week we collect, store, and analyze data for an average of 1,598,134 ($SD = 11,192$) apps. We enumerate all app URLs found on the Apple App Store and parse the privacy labels and metadata for each app. When parsing the data we extract: (i) app properties such as: (a) cost, (b) size, (c) developer, (d) content rating, (e) release date, (f) and genre, (ii) what data is collected, (iii) what is the purpose of the collection, (iv) and how the data will be used. Each run creates a snapshot of the App Store which we store in a *MySQL* database for analysis and comparison with past snapshots. The process proceeds in two stages: first, capturing the addition or removal of apps from the App Store; and second, creating a new snapshot for retrieving the associated app metadata and privacy label information.

Maintaining an accurate list of iOS apps. First, we retrieve and parse the site map XML file from <https://apps.apple.com>. While parsing the site map, the program extracts all of the individual app URLs on the Apple App store online. For example, <https://apps.apple.com/us/app/instagram/id389801252> is the app URL for the *Instagram* app. We then insert the URLs into a database table for use in retrieving the privacy label information and associated app metadata. A scheduled *cron* job on our server maintains an up-to-date list of all app URLs on the Apple App Store.

Retrieving privacy labels and app metadata. A process uses the full list of app URLs to retrieve the app privacy labels and metadata. Embedded inside each of the app web pages on the Apple app store is a JSON string that contains all of the app metadata associated with the app. The metadata, not including the privacy labels, contains 27 different pieces

of information about the app, such as the app name, version, size, type, user rating, genre, content rating, release date, seller name, and price. The privacy label information includes both the *Privacy Types* and *Data Categories*. The program reads and parses the JSON and stores the information in a *MySQL* relational database. Each row in the database table represents an app on the Apple App Store. To get the extended privacy label details, such as the *Purposes* and *Data Types*, for each app the program performs a GET request to the Apple catalog API at <https://amp-api.apps.apple.com/v1/catalog/>. The API response is a JSON object containing the extended privacy label information, which we parse and store in the database table linked to the associated app. To retrieve the full list of URLs, the Python script uses batch processing, with a batch size of 100 apps. Occasionally an error occurs due to too many requests, and we use an exponentially increasing back-off between retries. Any app that fails load after four retry attempts is recorded for a retry again after batch processing completes, allowing for HTTP status code 429 (Too Many Requests) to clear. The retrieval script runs weekly after the app list compilation is complete. The total retrieval time for all 1.6 million apps is between 72 to 96 hours with five processing nodes with independent IP addresses.

Ethical considerations We collect our privacy label data set using only publicly available Apple App Store web pages, and we do not abuse any protocols or hidden APIs to collect this data. Our measurement requests do incur an additional burden on Apple’s servers, and to alleviate that burden we place limits on the GET requests. We request the web pages containing the app metadata in batches containing only 100 apps and initiate a sleep time of 10 seconds between batches. We additionally deploy an exponential backoff for errors to further reduce over requesting Apple’s servers.

Resulting dataset. Since the composition of the $\sim 1.6\text{M}$ apps on the App Store changes each week, we gathered a total of 2,005,552 unique apps during our collection period. This set comprises both apps that existed at the beginning of our collection period and new apps

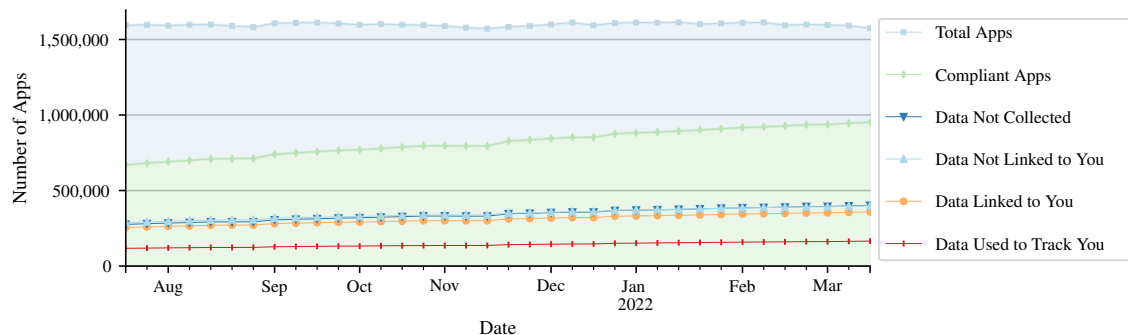


Figure 5.4: A longitudinal view over the 36-week collection period of the total number of apps and the total number of apps with privacy labels (compliant apps). For comparison, we also display the four *Privacy Types* over the same period. Each data point represents a snapshot of the Apple App Store on that date.

that were added during our collection, along with apps that were removed from the App Store. We identify each app based on the *App ID*, a unique value assigned by Apple, along with a *Run ID*, that we assigned to each week of our collection; the combination of these two values gives us the state of the privacy label of an app at a specific point in time. During our analysis, we observed relevant subsets of the captured apps, extracted their privacy labels at specific points in time, and combined their states to report on trends across the App Store. In the sections that follow, we highlight the specifics of each subset before expanding on observed metrics.

5.3 Overall App Store Trends

We observed an average of 1,598,134 ($SD = 11,192$) total apps per week in the App Store during our 36-week collection period, and there were 37,450 ($SD = 17,086$) newly published and 38,033 ($SD = 13,827$) removed apps per week, on average. In total, 60.5% ($n = 952,922$) of apps at the end of the collection period had a privacy label. This is an 18.4% increase from the start where only 42.1% ($n = 670,547$) had labels. There was a modest but steady increase of 0.5% ($n = 48,068$) per week, on average. There still remains 39.5% ($n = 621,908$) of apps without privacy labels. (Please refer to Figure 5.4 for the state of the app store at each

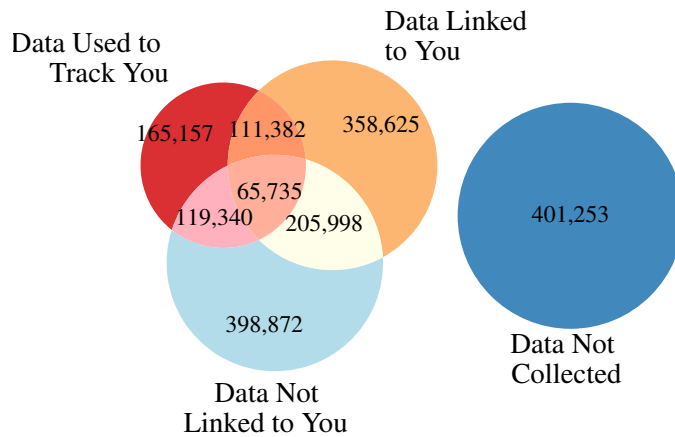


Figure 5.5: A Venn diagram of the number of apps in each of the four *Privacy Types*. *Data Not Linked To You* is mutually exclusive to the other three *Privacy Types*.

weekly snapshot.)

For the remainder of this section we will report only on apps with privacy labels ($n = 952,922$), so-called *compliant apps*. While the number of compliant apps steadily increased over the 36 weeks, so did the number of apps that have labels indicating that they track user data across apps and websites, which increased by 3,733 ($SD = 1,737$) apps on average each week for a total increase of 47,658 apps. The number of apps that link data to users' identities increased by 9,169 ($SD = 4,083$) apps on average each week for a total increase of 103,886 apps. But most new apps used the *Data Not Collected Privacy Type*. These increased on average 12,597 ($SD = 5,529$) apps per week for a total increase of 125,333 apps.

Relationship between privacy types. Considering the four *Privacy Types*, we find that 32.0% ($n = 305,250$) of apps have more than one *Privacy Type*, and *Data Not Collected*, which has the largest number of apps, is mutually exclusive with the three other *Privacy Types*. (See Figure 5.5.) We find that the three *Privacy Types* that signify data collection overlap with each other in varying amounts. For example, 6.9% ($n = 65,735$) of apps have all three *Privacy Types*. The most common ($n = 205,998$, 21.6%) combination of

Privacy Types is the *Data Linked to You* together with *Data Not Linked to You*. *Data Not Collected* is the most common *Privacy Type* with 42.1% ($n = 401,253$) of apps with labels. *Data Not Linked to You* is the next most common *Privacy Type* with 41.9% ($n = 398,872$) of apps with labels. These are followed by *Data Linked to You* with 37.6% ($n = 358,625$) and *Data Used to Track You* with 17.3% ($n = 165,157$) of apps with labels.

The purpose of data collection. The *Purposes* reported on the privacy label provide a compelling glimpse into why data is collected. *Purposes* are only provided for the *Data Linked to You* and *Data Not Linked to You* *Privacy Types*, as the purpose for the *Data Used to Track You* *Privacy Type* is by definition to track users for targeted advertising. Figure 5.6 shows the ratios of the six *Purposes*. The denominator for the ratios in this figure is the total number of apps labeled with the specific *Privacy Type*. In the final week of collection (March 17, 2022), the most common *Purposes* for collecting data linked to users' identities are *App Functionality* ($n = 314,205$; 87.6%) and *Analytics* ($n = 169,005$; 47.1%), and for collecting data not linked are *App Functionality* ($n = 283,134$; 71.0%) and *Analytics* ($n = 239,550$; 60.1%). These ratios of *Purpose* were shown to be stable over the study period. *App Functionality* was 88.0% and *Analytics* was 49.3% as purposes for collecting data linked to a users' identities. Moreover, *App Functionality* was 71.5% and *Analytics* was 62.2% as purposes for collecting data not linked to users in the first snapshot.

When we review the *Data Categories* contained under each *Purpose*, we find differences in the most common categories of data collection between *Data Linked to You* and *Data Not Linked to You*. For example, under *Data Linked to You* for the *Purpose* of *App Functionality*, *Contact Info* ($n = 248,040$; 69.2%) is the most common *Data Category*, but under *Data Not Linked to You* for the same *Purpose* of *App Functionality* it is *Diagnostics* ($n = 166,324$; 41.7%). This suggests that while *App Functionality* is often given as the reason for data collection, the category of data collected depends on whether that data is linked or not linked to a user's identity. A summary of the ratios of *Data Categories* by *Purpose* can be found

in the heatmaps of Figure C.3 and Figure C.4 in the Appendix.

Categories of data collected. Reviewing the *Data Categories* provides a view on what data is collected and how it is used. The most common *Data Categories* found under *Data Used to Track You* are *Identifiers* ($n = 113,641$; 68.8%) and *Usage Data* ($n = 98,098$, 59.4%). The *Identifiers* category contains the *User ID* and *Device ID Data Types*, which are frequently used to track users for the placement of targeted advertisements. Also frequently found under *Data Used to Track You* are *Diagnostics* ($n = 45,675$, 27.6%) and *Location* ($n = 44,162$, 26.7%). A small number of apps state that they collect very sensitive data for tracking purposes such as *Health & Fitness* ($n = 255$, 0.2%) and *Sensitive Info* ($n = 568$, 0.3%).

The most common *Data Categories* under *Data Linked to You* are *Contact Info* ($n = 261,712$; 73.0%) and *Identifiers* ($n = 233,536$; 65.1%). Apps also collected and linked very sensitive user data such as *Health & Fitness* ($n = 16,294$, 4.5%) and *Sensitive Info* ($n = 11,523$, 3.2%).

The most common under *Data Not Linked to You* are *Diagnostics* ($n = 259,769$; 65.1%) and *Usage Data* ($n = 198,608$; 49.8%). Figure 5.7 shows the ratios of the 14 *Data Categories* for each of the three *Privacy Types*.

Apps have 2.3 ($SD = 1.6$) *Data Categories* listed under *Data Used to Track You*, 6.2 ($SD = 6.2$) under *Data Linked to You*, and 3.6 ($SD = 3.5$) under *Data Not Linked to You*, on average. Viewed another way, there were a total of 4,022,272 instances of collected *Data Categories* on the app store, and 9.3% ($n = 373,741$) of them are *Data Used to Track You*, 54.9% ($n = 2,207,368$) are *Data Linked to You*, and 35.8% ($n = 1,441,163$) are *Data Not Linked to You*. A majority of data collected on the App Store is linked to users' identities and not anonymously collected.

Types of data collected. Finally, at the most detailed level are the *Data Types*. The specific detail about the type of data collected is important for broadly defined *Data Categories* like

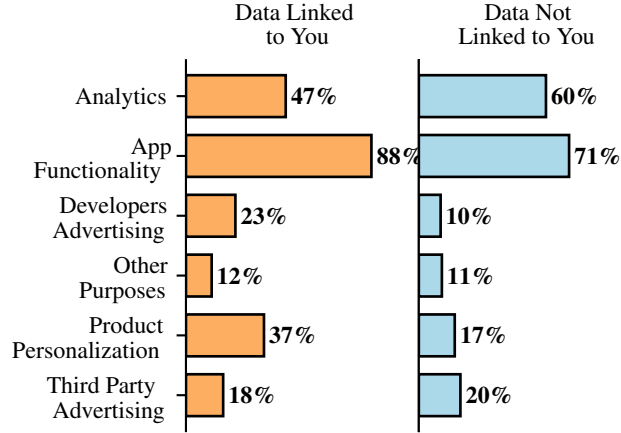


Figure 5.6: The ratios of the six *Purposes* for the *Data Linked to You* and *Data Not Linked to You Privacy Types*. The denominator is the number of apps in the specific *Privacy Type*.

Contact Info and *User Content*. For example, the *Contact Info* category may include the *Physical Address*, *Email Address*, *Name*, and *Phone Number Data Types*. For *Data Used to Track You* the most common *Data Types* collected are the *Device ID* ($n = 97,844$, 59.2%) followed by *Advertising Data* ($n = 73,951$, 44.8%) and *Product Interactions* ($n = 61,733$, 37.4%). For *Data Linked to You* the most common are *Email Address* ($n = 227,339$, 63.4%), *Name* ($n = 216,535$, 60.4%), *User ID* ($n = 189,144$, 52.7%), and *Phone Number* ($n = 179,049$, 49.9%). Refer to Figure 5.8 for full results.

5.4 Comparing App Metadata and Privacy Labels

In this section, we investigate how the metadata attributes of apps, such as content rating, app size, and price, relate to data collection practices as self-reported in the privacy labels, particularly the *Privacy Types* and *Data Categories*. Unless otherwise noted, all results are from the last weekly snapshot (Week 36, retrieved March 17, 2022).

Content rating. The content rating attribute is set by developers to communicate about the age appropriateness of an app, e.g, is it for adults or children. These content warnings are designated through Apple’s guidelines [8] and also align with parental control features,

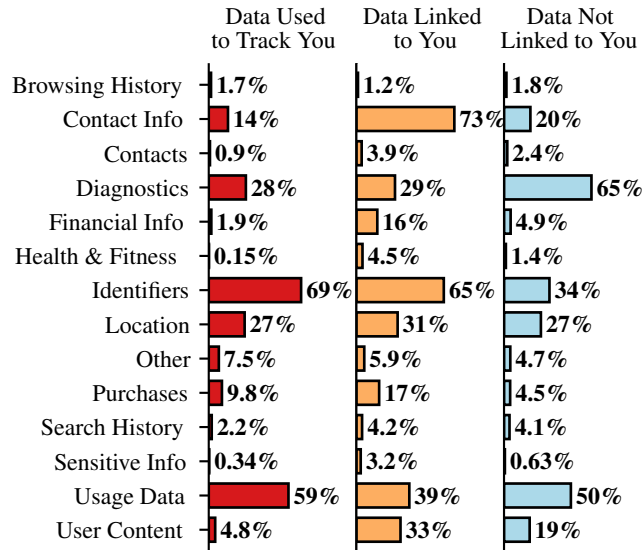


Figure 5.7: The ratios of the 14 *Data Categories* for each of three *Privacy Types*. The denominator is the number of apps in the specific *Privacy Type*.

which may disallow the usage of an app if the user is not of an appropriate age. More importantly, these content ratings also should comply with local policy requirements in each territory where the app is available; for example, this would be the Children’s Online Privacy Protection Act (COPPA) [25] in the US.

The choice of content rating age ranges are 4+, 9+, 12+, and 17+. Most apps with privacy labels have a content rating of ages 4+ ($n = 723,460$; 75.9%), while only 13.4% ($n = 127,924$) of apps with labels have a content rating of ages 17+. Of apps with privacy labels, 14.9% ($n = 107,864$) have a 4+ content-rating and are labeled with *Data Used to Track You*, and 34.7% ($n = 251,051$) are labeled with *Data Linked to You*. For the 17+ content rating, 18.4% ($n = 23,543$) were labeled with *Data Used to Track You* and 45.2% ($n = 57,823$) with *Data Linked to You*. Refer to Figure 5.4 for full details.

Apps with ratings of 4+ or 9+ designed for children would be subject to COPPA compliance in the US and require consent from parents to collect data that tracks minors under the age of 13. We did not perform a manual review of these apps to determine if they are actually directed at children, but the fact that so many apps with a 4+ or 9+ content

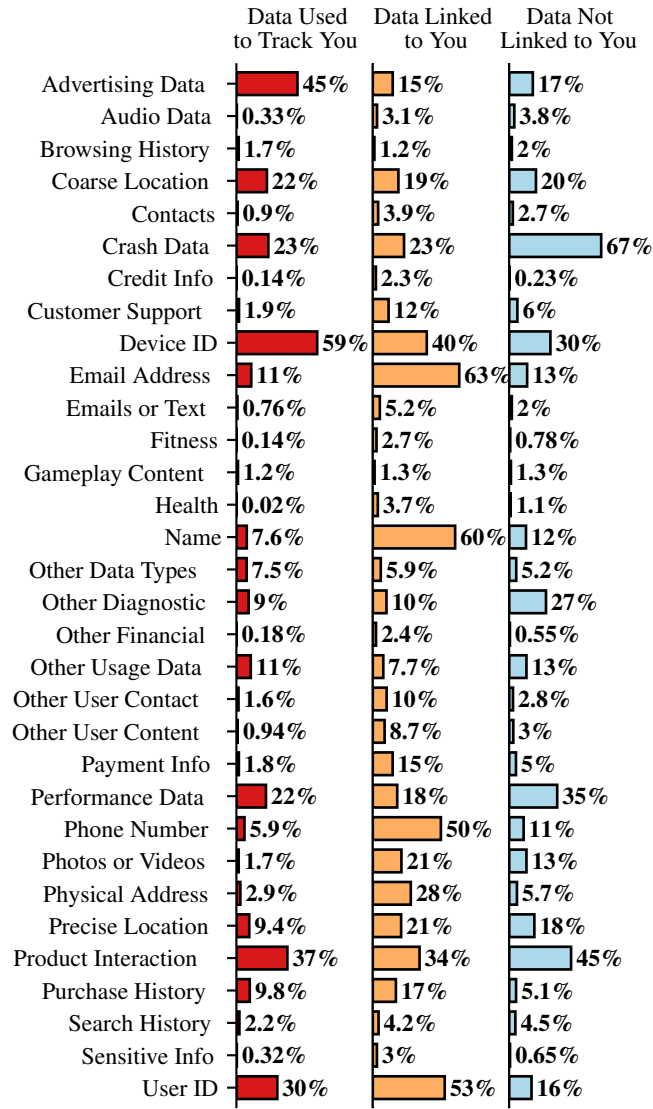


Figure 5.8: The ratios of the 32 *Data Types* for each of three *Privacy Types*. The denominator is the number of apps in the specific *Privacy Type*.

rating track data and would be available to children under the content rating guidelines (such as those used by parental controls) is problematic and worthy of further investigation.

Rating count. Users can review and rate apps on a five-star scale in the App Store. (Note that a user can rate an app without leaving a review.) The total number of app ratings offers a reasonable proxy for the overall popularity of an app, and we found that the higher the number of ratings for an app the more likely the app reports tracking user data and linking

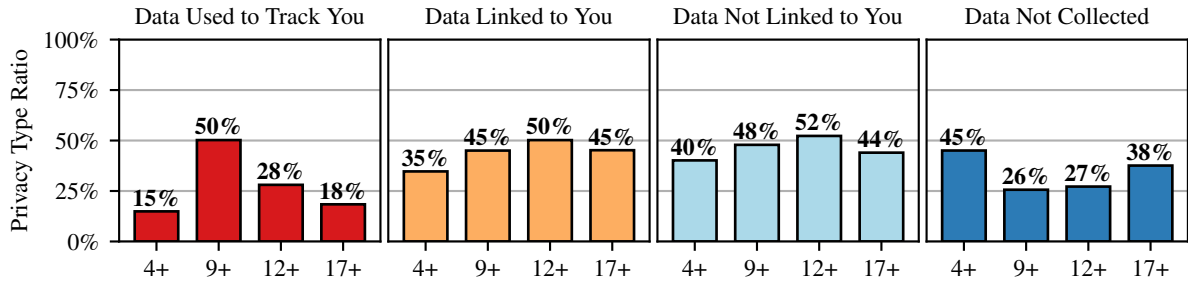


Figure 5.9: The ratios of content ratings for each of the four *Privacy Types*. The denominator is the number of apps with the designated content rating that have a privacy label.

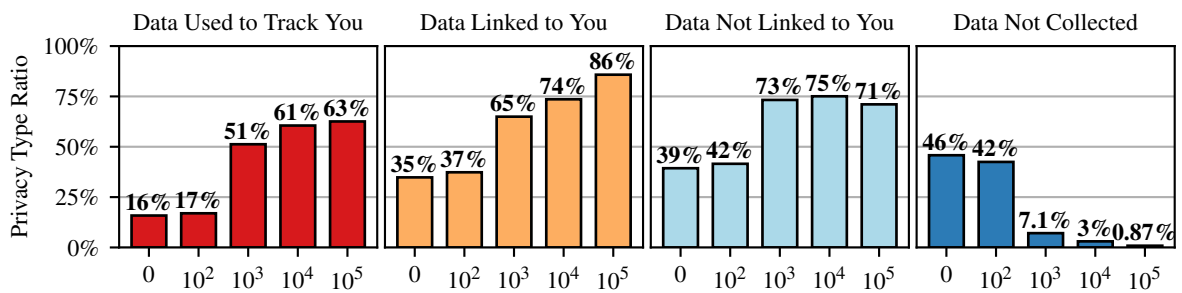


Figure 5.10: The ratios of the rating counts for each of the four *Privacy Types*. The denominator is the number of apps with the designated rating counts that have a privacy label. Apps with a larger number of user ratings are more likely to collect data, including data used to track users. Ratings counts are not localized metadata and apps with low ratings counts in the US region may have higher counts elsewhere.

data to users' identities, as reported in their privacy label. This may suggest that apps with more users are monetizing their popularity via surveillance surpluses [160], data collection beyond functionality and product improvement. That data may then be monetized for the purpose of targeted advertising and data sales.

Apps with over 100,000 ratings collect on average 19.79 ($SD = 12.06$, $M = 18$) *Data Categories* per app, while apps with fewer than 100,000 ratings only collect on average 7.28 ($SD = 6.84$, $M = 5$) *Data Categories* per app. Furthermore, 62.6% ($n = 361$) of apps with over 100,000 ratings were labeled with *Data Used to Track You* and 85.8% ($n = 495$) with *Data Linked to You*. This is a noticeably higher proportion of data collection for tracking and linking purposes than apps with fewer ratings. For comparison, apps with between 100

and 1,000 ratings, only 17.0% ($n = 160,046$) are labeled with *Data Used to Track You* and 37.3% ($n = 352,186$) with *Data Linked to You*. Refer to Figure 5.4 for full details.

Release date. The number of apps released in 2021 was 353,962, and due to Apple's requirements 100% of those apps have privacy labels. The apps released in 2021 make up 37.1% of all apps with labels. Newly released apps are the largest source of privacy-label compliant apps. Of the apps released in 2021, 16.3% were labeled with *Data Used to Track You* and 35.8% with *Data Linked to You*.

Only 45.8% ($n = 524,958$) of apps released before the December 8, 2020 were privacy-label compliant at the end of our collection window. This may suggest that a large number of apps are no longer being supported with developer updates, or there is a lack of incentive for developers to make privacy label updates without a requirement to do so. Refer to Figure C.1 in the Appendix for a detailed look at both the number of apps with privacy labels by release year and the number of those that have each *Privacy Type*.

App size. Apps that are larger in size collect more data used to track users and more data linked to users' identities. This may be due to the fact that apps with larger footprints contain additional software libraries that collect user data and track users. Additionally, game apps (as discussed below) are commonly tracking users and tend to require large downloads for graphics and other features.

Of apps with labels and a download size of greater than 1GB, 40.6% ($n = 1,737$) are labeled with *Data Used to Track You* and 35.8% ($n = 1,530$) are labeled with *Data Linked to You*. Compared with apps with labels and a download size of less than 100kB, only 1.2% ($n = 80$) are labeled with *Data Used to Track You* and 9.2% ($n = 588$) are labeled with *Data Linked to You*. Refer to Figure 5.4 in the Appendix for full details. Additionally, apps sized greater than 100MB collect on average 9.10 ($SD = 8.29$, $M = 7$) *Data Categories* per app, while apps sized less than 100MB only collect on average 6.68 ($SD = 6.18$, $M = 5$) *Data Categories* per app.

App price. Free apps with in-app purchases make up 13.3% ($n = 127,092$) of apps with labels, and of those apps, 42.1% ($n = 53,478$) are labeled with *Data Used to Track You*, 41.3% ($n = 52,481$) with *Data Linked to You*, and 60.7% ($n = 77,169$) with *Data Not Linked to You*. Free apps make up 82.4% ($n = 785,320$) of apps with labels, and of those apps 14.0% ($n = 110,316$) are labeled with *Data Used to Track You*, 38.7% ($n = 304,056$) with *Data Linked to You*, and 40.1% ($n = 315,211$) with *Data Not Linked to You*. Conversely, paid apps only account for 9.0% ($n = 85,941$) of compliant apps. Of those paid apps only 2.4% ($n = 885$) are labeled with *Data Used to Track You*, 4.2% ($n = 1,537$) with *Data Linked to You*, and 14.1% ($n = 5,199$) with *Data Not Linked to You*. Refer to Figure 5.4 for full details. Moreover, free apps collect on average 7.33 ($SD = 6.87$, $M = 5$) *Data Categories* per app, while paid apps only collect on average 4.15 ($SD = 4.54$, $M = 3$) *Data Categories* per app. This is in alignment with Scoccia et al. [134], who made this comparison on a small subset of apps. These findings differ from the work of Han et al. [70, 71] who investigated free and paid apps in the Android market based on inclusion of third-party advertising software, finding no differences between free and paid apps. These results, when looking at the privacy labels, suggest that free apps with in-app purchases are more likely to track users and link that data to users' identities, likely to generate revenues via targeted advertising and/or selling user data.

Top chart app genre. The Apple App Store offers categorization of apps into genres, such as Social Networking, Games, etc., and we analyzed the *Privacy Type* distributions for apps with labels in each of the genres. We found that Games apps are the most likely to track users, followed by Shopping, Music, Photo & Video, and Social Networking. Shopping apps are most likely to link data to users' identities, followed by Magazines & Newspapers, Finance, Lifestyle, Sports, and Social Networking apps. Shopping apps are the most likely to collect data not linked to users' identities followed by Magazines & Newspapers, Entertainment, Weather, and Lifestyle. Refer to Figure C.2 in the Appendix for full details.

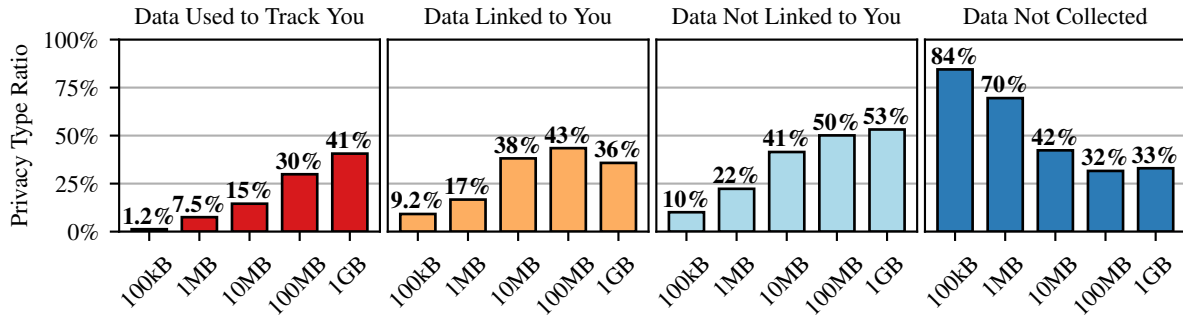


Figure 5.11: The ratios of app sizes for each of the four *Privacy Types*. The denominator is the number of apps with the designated app size that have a privacy label. Apps that are larger in size are more likely to collect data, including data used to track and linked to users.

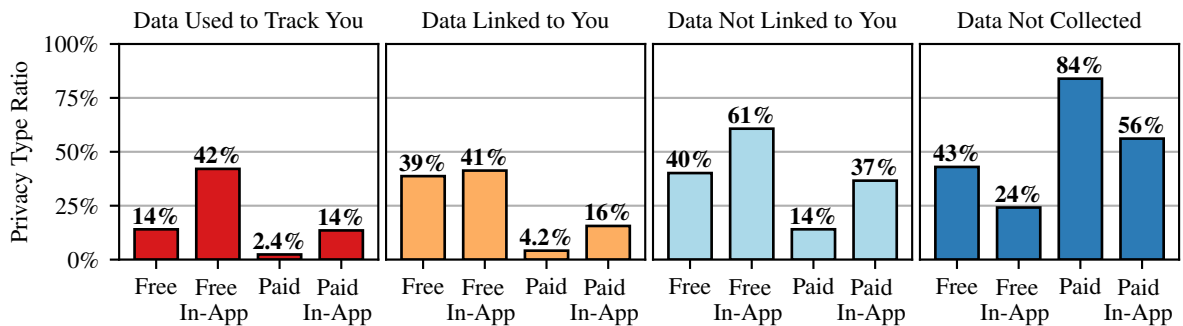


Figure 5.12: The ratios of app costs for each of the four *Privacy Types*. The denominator is the number of apps with the designated app cost that have a privacy label. Free apps are more likely than paid apps to collect data, including data used to track and linked to users.

Location data. The collection and use of location data is a source of concern for users of mobile apps [4]. We found that the *Location Data Category* was reported as *Data Used to Track You* by 44,162 apps, as *Data Linked to You* by 110,177 apps, and as *Data Not Linked to You* by 109,526 apps. In total, 217,582 apps collect location data, however this is only 22.8% of the total number of apps with privacy labels. Kollnig et. al. [87] found that 49.2% of iOS apps request opt-in permissions for access to device location data. This suggests that either the apps that request access to location data are keeping the data local to the device, or that location data collection is under-reported on the App Store privacy labels.

The most common *Purposes* given for collecting location data as *Data Linked to*

You are *App Functionality* ($n = 80,125$), *Analytics* ($n = 42,541$), *Product Personalization* ($n = 31,723$), and *Third-Party Advertising* ($n = 24,251$). The most common *Purposes* given for collecting location data in *Data Not Linked to You* are again *App Functionality* ($n = 76,718$), *Analytics* ($n = 27,555$), *Product Personalization* ($n = 26,126$), and *Third-Party Advertising* ($n = 14,700$). When we break the *Location* category down into the more detailed *Data Types* we find the *Coarse Location* is used by 36,940 apps, and *Precise Location* is used by 15,582 apps to track users across other apps and websites owned by third-parties. We also see that *Coarse Location* is used by 67,427 apps, and *Precise Location* is used by 74,524 apps, and this location data is collected and linked to users' identities. Finally, when we look at location data collected but not linked, we find that *Coarse Location* is used by 70,440 apps, and *Precise Location* is used by 66,185 apps.

5.5 Privacy Label Adoption and Changes

In this section we explore the ways in which apps come into compliance with the privacy label policy of the App Store, and how they change their privacy labels over time. Recall that adding privacy labels is mandatory for new apps that are added to the store following December of 2020 and for any app that submits an update. However, pre-existing apps can also update their privacy labels without updating the app (i.e., increasing the version number) by simply submitting a revision to their App Store page to include labels.

Voluntary vs. forced adoption of privacy labels. We first explore if there are differences in how developers choose to apply privacy labels for the first time because they have to, e.g., due to a version update or uploading a new app, or because they voluntarily added labels without changing other metadata. Figure 5.13 presents the differences in how these two types of apps are labeled, those without a version update (voluntary addition of privacy labels) to those with a version update (forced adoption of privacy labels). There are markedly different distributions in privacy label types. A much larger share of apps that were forced

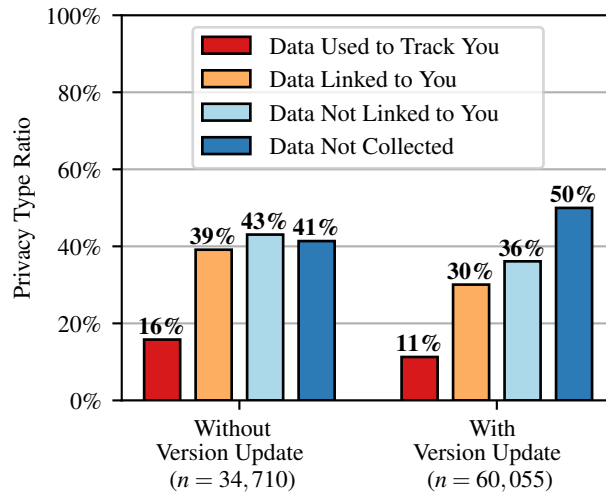


Figure 5.13: The changes in *Privacy Types* observed during the adoption of privacy labels by existing apps. App store policy enforced the addition of a privacy label to existing apps with their next version update. Apps that voluntarily added a label, without first releasing a version update, included more details about their data collection practices, while 50% of the apps that added a label with a version update stated that they did not collect any data.

to apply a privacy label choose *Data Not Collected* as a label (50% vs. 41%). A similar trend is observed for other types.

These results could suggest that developers voluntarily opting into privacy labels may be taking a more genuine approach to selecting labels as they were not required to do so. In contrast, developers that were forced to add labels may have thought of the process as onerous and simply an obstacle to the end goal of adding or updating an app. They simply selected labels for the purpose of expediting the process. The divergence in distributions may imply that many of the labels are in fact speculative and applied conveniently for a large share of the apps in the store, particularly given that many of the apps that have privacy labels are new apps, as opposed to version updates or voluntary updates. Moreover, the first labels that are applied are critical; we do not find that many apps (only 13,785) made changes to the privacy label during our 36-week observation period (more details below). The initial privacy labels persist, and so the accuracy of the first labeling is key.

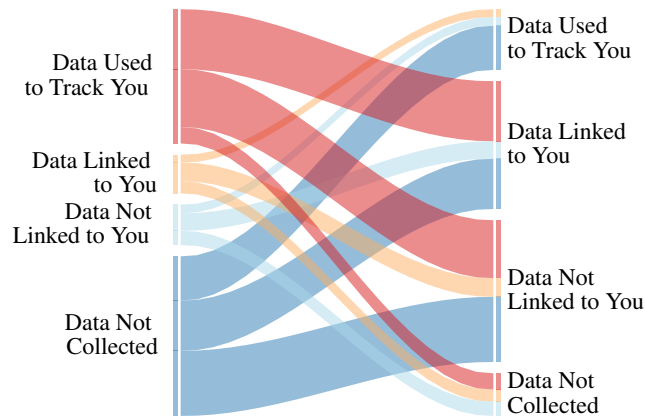


Figure 5.14: Changes to the *Privacy Types* associated with a *posture* during our crawl. While the most significant shifts were *from Data Not Collected*, the highest shifts were observed *to Data Linked to You* and *Data Not Linked to You*.

Lack of privacy label shifts. Of the 2,005,552 apps observed with a label during the 36-week collection period, only 0.007% ($n = 13,785$) apps added updates to their *Privacy Types*. That is, these apps were observed with one set of *Privacy Types* and at some point these *Privacy Types* changed in some way.

When apps do change their privacy labels there are two main shifts, as presented in Figure 5.14. The largest of these is a privacy label shift from *Data Not Collected* to any of the other three *Privacy Types* ($n = 10,788$). Most frequently this occurs from *Data Not Collected* to *Data Not Linked to You* ($n = 4,398$), followed by *Data Linked to You* ($n = 3,393$), and lastly, *Data Used to Track You* ($n = 2,984$), suggesting that when an app developer decides to change a privacy label (which is rare) when they originally had *Data Not Collected*, they are more likely to choose the less invasive of the *Privacy Types*, e.g., *Data Not Linked to You*. In contrast, the second most common shift is moving from a more invasive label (*Data Used to Track You*) to a less invasive one. *Data Used to Track You* ($n = 9,062$) shifted towards *Data Linked to You* ($n = 3,907$) and *Data Not Linked to You* ($n = 4,054$), and a small number ($n = 1,101$) moved from *Data Used to Track You* to *Data Not Collected*. Overall, very few shifts ($n = 2,885$) involved changes from a *Privacy Type* where

Table 5.1: Changes observed in the addition and removal of data categories from privacy labels. Twice as many apps that changed the composition of their labels added a data category as those that removed a data category.

Data Category	Added	Removed
Identifiers	5,556	1,853
Usage Data	5,079	1,458
Diagnostics	4,258	1,280
Contact Info	3,442	1,322
Location	3,535	1,408
User Content	2,364	1,081
Purchases	1,158	500
Other	1,115	589
Financial Info	1,077	450
Search History	615	458
Contacts	429	195
Browsing History	322	354
Health and Fitness	258	106
Sensitive Info	216	110

data is collected and/or tracked to *Data Not Collected*. These patterns are also observed in Figure C.5 in the Appendix, where shifts between each week are graphed.

Privacy Label shifting with version vs. metadata updates. Apps that shift privacy labels can do so in two ways. Either the developer can change a privacy label due to a version update to the app, or they can simply update the metadata of the app to change the privacy label without a version update. We split the 13,785 apps that performed voluntary updates to their *Privacy Types* into these two categories, and the results are presented in Table C.1 in the Appendix. There were no observable difference, suggesting that voluntary updates to privacy labels are consistent in these two cases.

Shifts in data categories. We also measured how the data categories and their associated *Privacy Types* changed for apps that had privacy labels. Recall that an app’s privacy label is multi-leveled, beginning with the *Privacy Type* (e.g., *Data Not Collected* or *Data Used to Track You*), and then under that, a developer can note the specific data categories

being collected/tracked, as well as the purposes. Importantly, the *Data Categories* and *Purposes* can be changed without necessarily changing the *Privacy Types*.

When we observe shifts in *Data Categories*, developers almost always added new categories rather than removing existing ones. The most common *Data Categories* added and removed are presented in Table 5.1. The most commonly added categories are *Identifiers*, *Usage Data*, and *Diagnostics*, but these were also the most commonly removed *Data Categories*.

Data Categories can also be moved between *Privacy Types*; for example, *Contact Info* may have previously been of the type *Data Linked to You* and is now *Data Used to Track You*. Similar to *Data Categories* that were added, *Data Categories* that were shifted most often between *Privacy Types* were *Usage Data*, *Identifiers*, and *Diagnostics*. Specific to these three *Data Categories*, the most prominent shifts are towards *Data Used to Track You* and *Data Linked to You* from *Data Not Linked to You*. This is in contrast to other *Data Categories*, where we more commonly observed a shift from *Data Linked to You* to *Data Not Linked to You*. It is unclear if these shifts indicate that developers are making their *Data Categories* more accurate or are attempting to obscure data collection practices of more sensitive data. That is, they are willing to note that data is collected, but do not want to indicate that it is actually linked to users.

Shifts by app genre. We analyzed apps based on their App Store assigned *genres* to determine if that impacted shifts in *Data Categories* between *Privacy Types*. We found that *Games* engage in multiple *Data Categories* shifts, more so than any other *genres*. The *Data Categories* that are most often shifted are *Usage Data* ($n = 733$), *Identifiers* ($n = 565$), *Diagnostics* ($n = 239$), *Location* ($n = 205$), *User Content* ($n = 48$), *Purchases* ($n = 77$), and *Other* ($n = 46$) *Data Categories* between *Privacy Types*. The audience of mobile app game users is “expansive” [33, 131] and is a lucrative market for in-app advertising, which may explain why so many *Data Categories* shifts occur [36]. Refer to Table C.2 in the Appendix

for detailed counts of *Privacy Type* shifts associated with *Data Categories*, along with the most prominent *genre* that participates in these shifts.

Shifts in purposes. *Data Categories* can also be assigned different *Purposes* when labels change; for example, a label that had previously shown that *Identifiers* were collected for the purpose of *App Functionality* may now be changed during an update to state the purpose as *Third Party Advertising*. Across different *Data Categories*, labels most often change (or add) *Purposes* to *App Functionality*, *Analytics*, and *Product Personalization*. These shifts potentially indicate developers attempting to associate data collection practices with more beneficial purposes. More details can be found in Table C.3 in the Appendix, which includes detailed counts of *Purpose* shifts associated with *Data Categories*.

5.6 Discussion and Conclusion

This paper presents a large-scale, longitudinal evaluation of privacy labels in the Apple App Store. To accomplish this we collected and analyzed the privacy labels and other metadata for 1.6 million apps every week for 36 weeks. We identified a steady increase in the number of apps with privacy labels and the potential for under-reporting of data collection by apps that were forced to provide a label on a version update as compared to apps that were labeled voluntarily. We find that the data collected is most often linked to users' identities for the purposes of app functionality and analytics. Finally, we discovered that certain app attributes, such as free apps, apps with a higher number of ratings, and apps that are larger in size, indicate a greater number of categories of data collection. In the rest of this section we discuss the implications of these findings and areas for future research.

Prevalence of tracking. The most invasive privacy label is *Data Used to Track You*, and when an app has this label, it often includes *Data Categories* that would cover device identifiers, contact information, name and/or email address, as we detail in Section 5.3. The

data harvested by these apps for tracking could potentially be used by the app developers to “connect the dots” about user behavior by merging this info with other user activity across apps, websites, and purchases in physical stores. This type of tracking likely remains prevalent despite efforts to curtail it. The advent of the App Tracking Transparency [9] framework in iOS 14.5 in April of 2021 is an attempt to limit this effect and requires apps to request user authorization to access app-related data for tracking. So far, it appears effective to an extent, but some developers found ways to circumvent its restrictions [89].

The high cost of free apps. Many free apps are only free because they partake in the lucrative collection and sharing of personal data. As we presented in Section 5.4, free apps collect from, on average, three more categories of data than paid apps. When you combine this with other app metadata properties the contrast is even more stark. For example, free apps that also have more than 100,000 ratings (an indicator of audience size) collect on average 19.82 ($SD = 12.05$, $M = 18$) *Data Categories* per app. The large audience increases the size of surveillance surplus, which in turn may make it harder for app sellers to resist collecting a wider range of data to increase profits. For instance, when we consider *Social Networking* as the app genre combined with free apps that have more than 100,000 ratings we find they collect an average of 25.85 ($SD = 17.75$, $M = 23$) *Data Categories* per app, further exacerbating privacy issues by expanding individual collection to their contacts on the platform as well.

Our findings differ somewhat from prior work by Han et al. [70, 71], where they used the inclusion of third-party libraries as a proxy for privacy behaviors, and compared free apps with their paid counterparts. While they found no clear difference in privacy behaviors between these pairs, our analysis looks at free and paid apps from an ecosystem-wide perspective. At the very least, our findings confirm that the popular free *Games*, *Shopping*, and *Social Networking* apps are also the top collectors of user data.

Empowering users. Ideally, Apple iOS users would compare the privacy labels of apps with similar functionalities and select the app that best satisfies their personal privacy preferences and expectations. However, there is evidence from prior work that users may not make such choices using install-time information. In many ways, privacy labels function similarly to install-time permissions from Android [50]. They are a one-time opportunity, prior to installation, for the user to review and consider the privacy implications before installing an application. And like install-time permissions, users are likely not sufficiently informed nor sufficiently motivated to take action as a result of the privacy labels.

A second challenge is that the privacy labels are integrated into the App Store, not the iOS device itself. There are no mechanisms for users to review already installed apps, other than go to the App Store and select each app one by one. And even if a user were to perform this action, there is no mechanism for them to become aware of changes or updates in the privacy labels for apps over time. It remains unclear how and where privacy labels are intended to assist users in making informed decisions about their apps.

A question of accuracy. There have been news reports [55] and research [89] about the inaccuracies found in App Store privacy labels. We illustrate in Section 5.5 that when apps are forced to add privacy labels due to a version change or as a new app they are more frequently providing the *Data Not Collected Privacy Type*, potentially out of convenience, expediency, or deception. Furthermore, we describe in Section 5.4 how 22.8% of apps with privacy labels report collecting location data, but research [87] has found that 49.2% of apps request access to location data. These discoveries suggest under-reporting of data collection via privacy labels. And anecdotally, when reviewing a few apps manually, we observed apps that have a *Data Not Collected* label that conflicts with data collection practices outlined in their privacy policy. Future research is required to determine the full scope and nature of the inaccuracies in the privacy label ecosystem.

Without additional oversight from Apple and without negative consequences for inaccu-

rate privacy labels, developers may simply never be motivated to create labels that accurately reflect their apps' data collection practices. A lack of credibility and integrity in the privacy label model will ultimately erode user confidence in the system and reduce the chances that market forces will force app developers to make more privacy preserving applications going forward. Inaccurate labels can also be harmful to users when users' privacy expectations are misaligned with the true nature of an app's data collection practices.

Chapter 6: Proposed Work and Timeline

This thesis proposal describes the study of privacy disclosure mechanisms that notify users of what information will be collected, how it will be used, and with whom it will be shared. App-based privacy nutrition labels are an example of such a disclosure mechanism provided to users in a standardized and simplified format. To better understand the impact of app-based privacy labels on users' perceptions of risk and willingness to install an app we propose to conduct a large scale user perceptions study. The details regarding this study along with an analysis plan and timeline are provided in section 6.1.

This thesis proposal also describes the study of privacy control mechanisms that allow users to manage their information: limiting the use information or sharing with third-parties, and requesting removal or modification of information. The Google Ad Settings dashboard is an example of such a control mechanism that allows users to modify the way in which Google stores ad inference information, by limiting the types of interest that can be assigned, removing existing interests, or opting out of advertising inferences all together. To measure how interacting with Google ad interests changes the way in which users think about the benefits and concerns of data collection we propose to conduct a pre/post intervention study. The details regarding this study along with an analysis plan and timeline are provided in section 6.2.

6.1 The Impact of App-Based Privacy Labels on Users

Privacy labels have the potential to help users make informed choices when selecting an application to install. But it is unknown if privacy labels have the ability to convey privacy risk to users and what impact labels have on users' willingness to install an application. It is also unknown if users care about privacy risks even if they are conveyed by the privacy labels. Therefore, it is important to understand whether privacy labels lead to better privacy

outcomes for users such that users' privacy expectations align with the actual behavior of the apps that they use.

Impact on User Behavior. This study will research user perceptions, understanding, and behavior, with regards to the application privacy information (app privacy labels) added to the Apple iPhone App Store in December 2020. The study will consist of a survey that will ask iPhone users directly about their experiences with the privacy labels on the Apple App Store, and how these privacy labels impacted their app purchase decision making.

Apple says there are now more than 1 billion active iPhones [81]. In December 2020, Apple introduced privacy nutrition labels to their app listings in the iPhone App Store. The labels serve as an alternative to lengthy and difficult to read privacy policies that research has shown are never read by users. The new privacy nutrition labels could help iPhone users consider privacy and security when selecting an iPhone app and help them make more informed decisions about which apps to install.

Research Questions. We are going to investigate the following research questions related to app privacy labels:

RQ1 How do users understand the data collection information summarized on the privacy label?

RQ2 Which app privacy label attributes significantly influence user risk perception and willingness to install and in what ways?

RQ3 Which privacy labels are the most concerning to users and do users' perception of privacy risk align with actual risk?

These results will speak directly to a growing body of results regarding privacy communication to end-users of mobile applications, and how this information is employed by users to protect themselves from personal data collection. We have an opportunity to understand

user behavior in regards to privacy communication provided in a new format, namely the privacy label, and its impact on whether or not to install an application on their device. This will provide an important baseline that can be compared to prior research on the topic, and then used as a reference in future research.

The full survey questionnaire can be found in Appendix E.1. The procedure for the online survey will be as follows:

1. We will recruit participants using the online survey participant recruitment service Prolific.
2. Participants will be directed to the online survey platform Qualtrics where we will inquire about their perceptions and understanding of Apple App Store privacy labels.
3. Participants will be informed about the procedures and provide consent before proceeding.
4. The survey should take between 10 minutes to 30 minutes to complete.
5. The survey will inquire about participants' perceptions of app privacy labels and how they would modify their app installation behavior based on the information displayed in an app privacy label.
6. Participants will be shown various Apple iPhone apps along with their associated app privacy labels and asked to share their level of concern and their reasons for concern if any.
7. Participants will be shown various app privacy labels independent of iPhone apps, and will be asked to describe their risk perceptions, their confidence in the meaning of the privacy label, and their willingness to install an iPhone app with that label.
8. Participants may opt-out of the survey at any time by clicking the opt-out button.

Analysis Plan. The results from the online survey will be analyzed with the following statistical tests, thematic analysis, and the factors to be compared:

1. We will perform Wilcoxon signed-rank tests for repeated measurements on the Likert responses.
2. We will create ordinal regression models to describe the impact of privacy labels on risk perception and willingness to install an app after viewing the app's privacy labels.
3. We will conduct qualitative open coding to analyze free-response questions. A primary coder from the research team will craft a codebook and identify descriptive themes by coding each question. A secondary coder will code a 20% sub-sample from each of the free-response questions over several rounds providing feedback on the codebook and iterating with the primary coder until inter-coder agreement is reached (Cohen's $\kappa > 0.7$).

The detailed analysis plan can be found in Table 6.1.

Table 6.1: Detailed Analysis Plan.

RQ#	Question	Analysis Type	Description
RQ1	Q8	Quantitative	Create a Likert confidence scale bar chart for each privacy label that reports the participants' knowledge of what the label means.
RQ2	Q6	Quantitative	Create a bar chart for each app that reports whether the participant ever considered installing the app.
RQ2	Q7	Qualitative	Perform qualitative open coding to analyze free-responses to find themes that detail the participants' explanations about why they decided not to install the app.
RQ2	Q11	Quantitative	Create a bar chart for each app that reports whether the privacy label alters the participants' willingness to install the app.
RQ2	Q12	Qualitative	Perform qualitative open coding to analyze free-responses to find themes that detail the participants' explanations about why the privacy label alters their willingness to install the app.
RQ2	Q9	Regression	Create a Cumulative Link Mixed Model (CLMM) to describe the perception of privacy risk from the label after viewing the app and privacy label information. Dependent variable: perception of privacy risk from the label. Independent variables: willingness to install the app, privacy label type, type of app, have installed or considered installing the app.

Table 6.2: Detailed Analysis Plan Continued.

RQ#	Question	Analysis Type	Description
RQ3	Q1	Quantitative	Create a Likert concern scale bar chart for each app that reports the participants' concern for how the app will collect, store, and use information.
RQ3	Q2	Qualitative	Perform qualitative open coding to analyze free-responses to find themes that detail the participants' explanations about why their concern for how the app will collect, store, and use information.
RQ3	Q9	Quantitative	Create a Likert risk scale bar chart for each privacy label that reports the participants' perception of whether the privacy label decreases, increases, does not have any impact on the privacy and security risks associated with a specific app.
RQ3	Q10	Qualitative	Perform qualitative open coding to analyze free-responses to find themes that detail the participants' explanations about why they believe the privacy label decreases, increases, does not have any impact on the privacy and security risks associated with a specific app.

Table 6.3: Timeline for proposed tasks.

Task	Description	Status	Estimated Time	Target Date
Online Survey	Create an online survey to measure users' understanding, risk assessment, and impact on behavior of app-based privacy labels.	Completed		
IRB Approval	Obtain IRB approval for administering online survey to participants.	Completed		
Publish Survey	Publish our online survey on Prolific and gather data from survey participants.	Not Started	Two Weeks	October 2022
Qualitative Coding	Qualitative coding of open text responses from online survey.	Not Started	Two Weeks	November 2022
Quantitative Analysis	Quantitative analysis of survey results including regression analysis, statistical analysis, and descriptive analysis.	Not Started	One Month	December 2022
Writing	Write chapter that details the results of the study.	Not Started	One Month	January 2022

6.2 Ad Settings Study

Google allows users of their services to review, manage, and disable ad settings. Ad settings are a list of user related interests which have been inferred by Google from a user's Google activity. The ad settings interface can provide users insight into what Google infers about a user from their past activities using Google products, and the data-collection involved to make these inferences. The goal of this study is to investigate the privacy concerns and attitudes of people before and after viewing their Google ad settings page in order to develop a better understanding of which privacy trade-offs users make to use an internet service and which privacy control mechanisms they wish to have.

These results will speak directly to a growing body of results regarding Google's data collection practices, how these practices fit into the strategies employed by users, and how the access to data from combined Google services increases privacy concerns. We have a unique opportunity to understand user reactions to the inferences that Google makes about them based on their own data collected by Google being presented to them directly in the study. This study provides an important baseline that can be compared to prior research on the topic, and then used as a reference in future research.

Research Questions. We are going to investigate the following research questions related to Google advertisement inferences:

RQ1 How do users believe inferences are made about them based on their activities?

RQ2 How do users benefit and concern from Google's data collection change when considering the inferences made about their activities?

RQ3 How do users react to and the accuracy of Google's inferences?

The full survey questionnaire can be found in Appendix D.1. The procedure for the online survey will be as follows:

1. We will recruit participants using the online survey participant recruitment service

Prolific.

2. Participants will be directed to an online survey where we will inquire about their perceptions and understanding of Google Ad Settings and the inferences made from their Google activities.
3. Participants will be informed about the procedures and provide consent before proceeding.
4. Participants will be provided with detailed instructions for the installation of a web browser extension. The browser extension works in concert with the survey and will present participants with example Ad Settings inferred by Google on their Ad Settings page and activities recorded by Google as part of their My Activity log.
5. The survey will then inquire about participants' privacy concerns and attitudes before and after viewing their Ad Settings inferences.
6. Then the survey will display and select activities from participants' Google My Activity page and ask participants to match the activity with a list of possible inferences for that activity.
7. Next, the survey will display the participants actual ad inferences and ask questions related to the ad inferences, such as the relevance and accuracy of the inference, and concerns and benefits.
8. After participants complete or withdraw from the survey, they will be provided with detailed instructions for the removal of the web browser extension. When the web browser extension is removed, local data stored on the participants' machines is also removed.

Analysis Plan. The results from the online survey will be analyzed with the following statistical tests, thematic analysis, and the factors to be compared:

1. We will perform Wilcoxon signed-rank tests for repeated measurements on the Likert responses to the pre and post-exposure questions on concern and benefit.
2. We will also perform proportional odds logistic regressions to analyze which factors, in addition to the intervention, that may have influenced the Likert responses moving up or down the scales for concern and benefit.
3. We will create binomial logistic regression and ordinal regression models to describe the level of concern and frequency of benefit after visiting the Google Ad settings page.
4. We will conduct qualitative open coding to analyze free-response questions. A primary coder from the research team will craft a codebook and identify descriptive themes by coding each question. A secondary coder will code a 20% sub-sample from each of the free-response questions over several rounds providing feedback on the codebook and iterating with the primary coder until inter-coder agreement is reached (Cohen's $\kappa > 0.7$).

The detailed analysis plan can be found in Table 6.4.

Table 6.4: Detailed Analysis Plan.

RQ#	Question	Analysis Type	Description
RQ1	I(1-9)	Descriptive	Describe what ad interests participants choose and why.
RQ1	I(1-9)	Quantitative	Compare how accurate participant interest choices are with the ground truth from their ad settings page interests.
RQ1	I(1-9)	Qualitative	Create buckets of interests that are categorized by type. Describe the type and number of interest groups.
RQ2	Q2/Q9	Quantitative	Create a Likert scale bar chart that compares participants' reported concern about data collection pre and post matching their online activities to inferred interests.
RQ2	Q3/Q10	Quantitative	Create a Likert scale bar chart that compares participants' reported benefit from data collection pre and post matching their online activities to inferred interests.
RQ2	Q2/Q9	Regression	Create an ordinal regression model to describe the level of concern after participants match their online activities to inferred interests. Dependent variable: post concern. Independent variables: pre concern, matching game accuracy, increasing benefit, gender, education, has STEM background.
RQ2	Q3/Q10	Regression	Create an ordinal regression model to describe the level of benefit after participants match their online activities to ad interests. Dependent variable: post benefit. Independent variables: pre benefit, ad interest matching accuracy, decreasing concern, gender, education, has STEM background.
RQ2	Q2_A/Q9_A	Qualitative	Perform qualitative open coding to analyze free-responses to find themes that detail the participants reasons for concern regarding data collection both pre and post matching online activities with inferred inferences.
RQ2	Q3_A/Q10_A	Qualitative	Perform qualitative open coding to analyze free-responses to find themes that detail the participants reasons for benefiting from data collection both pre and post matching online activities with inferred inferences.

Table 6.5: Detailed Analysis Plan Continued.

RQ#	Question	Analysis Type	Description
RQ3	Q11	Quantitative	Create a Likert scale bar chart that compares participants' reported concern about data collection after viewing the ad interests that Google infers about them.
RQ3	Q11_A	Qualitative	Perform qualitative open coding to analyze free-responses to find themes that detail the participants reasons for concern regarding data collection after viewing the ad interests that Google infers about them.
RQ3	Q12	Quantitative	Create a Likert scale bar chart that compares participants' reported benefit from data collection after viewing the ad interests that Google infers about them.
RQ3	Q12_A	Qualitative	Perform qualitative open coding to analyze free-responses to find themes that detail the participants reasons for benefiting from data collection after viewing the ad interests that Google infers about them.
RQ3	Q11	Quantitative	Create a Likert scale bar chart that describes the participants' reported perceptions about the accuracy of the ad interests inferred about them.
RQ3	Q12_A	Qualitative	Perform qualitative open coding to analyze free-responses to find themes that detail the participants explanations about why they perceived the inferred ad inferences to be accurate or not accurate.

Table 6.6: Timeline for proposed tasks.

Task	Description	Status	Estimated Time	Target Date
Online Survey	Create an online survey to measure users' understanding, concerns and benefits of advertising inferences.	Completed		
IRB Approval	Obtain IRB approval for administering online survey to participants.	Completed		
Publish Survey	Publish our online survey on Prolific and gather data from survey participants.	Completed		
Qualitative Coding	Qualitative coding of open text responses from online survey.	Completed		
Quantitative Analysis	Quantitative analysis of survey results including regression analysis, statistical analysis, and descriptive analysis.	In Progress	One Month	November 2022
Writing	Write chapter that details the results of the study.	Not Started	One Month	December 2022

Bibliography

- [1] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS '14*, page 674–689, New York, NY, USA, 2014. Association for Computing Machinery.
- [2] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Computing Surveys*, 50(3):44:1–44:41, August 2017.
- [3] Federal Trade Commission: Consumer Advice. How To Use the EnergyGuide Label To Shop for Home Appliances. <http://consumer.ftc.gov/articles/how-use-energyguide-label-shop-home-appliances>, May 2021.
- [4] Hazim Almuhiemedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. Your Location has been Shared 5,398 Times! A Field Study on Mobile App Privacy Nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, pages 787–796, New York, NY, USA, April 2015. Association for Computing Machinery.
- [5] Athanasios Andreou, Marcio Silva, Fabricio Benevenuto, Oana Goga, Patrick Loiseau, and Alan Mislove. Measuring the Facebook Advertising Ecosystem. In *Proceedings 2019 Network and Distributed System Security Symposium*, San Diego, CA, 2019. Internet Society.
- [6] Julio Angulo, Simone Fischer-Hübner, Tobias Pulls, and Erik Wästlund. Usable Transparency with the Data Track: A Tool for Visualizing Data Disclosures. In *ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '15*, pages 1803–1808, Seoul, Republic of Korea, April 2015. ACM.
- [7] Apple. Accessing private data - Patterns - Human Interface Guidelines - Design - Apple Developer. <https://developer.apple.com/design/human-interface-guidelines/ios/app-architecture/accessing-user-data>. Last Accessed: October 28, 2022.
- [8] Apple. App Store Review Guidelines - Apple Developer. <https://developer.apple.com/app-store/review/guidelines>. Last Accessed: October 28, 2022.
- [9] Apple. App Tracking Transparency | Apple Developer Documentation. <https://developer.apple.com/documentation/apptrackingtransparency>. Last Accessed: October 28, 2022.

- [10] Apple. Authentication Services | Apple Developer Documentation. <https://developer.apple.com/documentation/authenticationservices>. Last Accessed: October 28, 2022.
- [11] Apple. Control access to information in apps on iPhone - Apple Support. <https://support.apple.com/guide/iphone/control-access-to-information-in-apps-iph251e92810/ios>. Last Accessed: October 28, 2022.
- [12] Apple. Submissions now accepted through the holidays - Latest News - Apple Developer. <https://developer.apple.com/news/?id=y4fgrhhe>, November 2021.
- [13] Apple Developer. App privacy labels now live on the App Store - Latest News - Apple Developer.
- [14] Guangdong Bai, Jike Lei, Guozhu Meng, Sai Sathyanarayan Venkatraman, Prateek Saxena, Jun Sun, Yang Liu, and Jin Song Dong. Authscan: Automatic Extraction of Web Authentication Protocols From Implementations. In *Network & Distributed System Security Symposium*, 2013.
- [15] Rebecca Balebako, Florian Schaub, Idris Adjerid, Alessandro Acquisti, and Lorrie Cranor. The impact of timing on the salience of smartphone app privacy notices. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*, SPSM '15, page 63–74, New York, NY, USA, 2015. Association for Computing Machinery.
- [16] Natã M. Barbosa, Gang Wang, Blase Ur, and Yang Wang. Who Am I? A Design Probe Exploring Real-Time Transparency about Online and Offline User Profiling Underlying Targeted Ads. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3):88:1–88:32, September 2021.
- [17] Muhammad Ahmad Bashir, Umar Farooq, Maryam Shahid, Muhammad Fareed Zaffar, and Christo Wilson. Quantity vs. Quality: Evaluating User Interest Profiles Using Ad Preference Managers. In *Proceedings 2019 Network and Distributed System Security Symposium*, San Diego, CA, 2019. Internet Society.
- [18] Lujo Bauer, Cristian Bravo-Lillo, Elli Fragkaki, and William Melicher. A Comparison of Users' Perceptions of and Willingness to Use Google, Facebook, and Google+ Single-Sign-On Functionality. In *Workshop on Digital Identity Management*, 2013.
- [19] Nataliia Bielova. Web Tracking Technologies and Protection Mechanisms. In *ACM Conference on Computer and Communications Security, CCS '17*, pages 2607–2609, Dallas, Texas, USA, October 2017. ACM.
- [20] Abine Blur. Abine Blur: passwords, payments, & privacy. <https://www.abine.com>. Last Accessed: October 28, 2022.
- [21] Dieter Bohn. Google Is Sending a Complicated Privacy Email to Everyone. *The Verge*, August 2020. <https://www.theverge.com/2020/8/5/21354805>.

- [22] Farah Chanchary and Sonia Chiasson. User Perceptions of Sharing, Advertising, and Tracking. In *Symposium on Usable Privacy and Security*, SOUPS '15, pages 53–67, Ottawa, Canada, July 2015. USENIX.
- [23] Douglas Clark. Google's US Ad Revenues to Drop for the First Time. <https://newsroom.emarketer.com/newsroom/index.php/google-ad-revenues-to-drop-for-the-first-time>, June 2020.
- [24] European Commission. Energy-efficient products. https://ec.europa.eu/info/energy-climate-change-environment/standards-tools-and-labels/products-labelling-rules-and-requirements/energy-label-and-ecodesign/energy-efficient-products_en. Last Accessed: October 28, 2022.
- [25] Federal Trade Commission. Children's Online Privacy Protection Rule ("COPPA"). <http://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa>. Last Accessed: October 28, 2022.
- [26] OAuth Community. OAuth Community Site. <https://oauth.net>. Last Accessed: June 2, 2021.
- [27] Comscore, Inc. Unique U.S. Visitors Top 50 Platforms (Desktop and Mobile). <https://www.comscore.com/Insights/Rankings?country=US>. Last Accessed: October 28, 2022.
- [28] Kovila P.L. Coopamootoo. Usage Patterns of Privacy-Enhancing Technologies. In *ACM Conference on Computer and Communications Security*, CCS '20, pages 1371–1390, Virtual Conference, October 2020. ACM.
- [29] Lorrie Faith Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *J. on Telecomm. & High Tech. L.*, 10:273, 2012.
- [30] Lorrie Faith Cranor, Candice Hoke, Pedro Leon, and Alyssa Au. Are They Worth Reading? An In-Depth Analysis of Online Advertising Companies' Privacy Policies. SSRN Scholarly Paper ID 2418590, Social Science Research Network, Rochester, NY, March 2014.
- [31] Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1):92–112, March 2015.
- [32] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. In *Symposium on Network and Distributed System Security*, NDSS '19, San Diego, California, USA, February 2019. ISOC.
- [33] Deloitte. Global Mobile Consumer Survey | The state of the smartphone. <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/technology>

-media-telecommunications/deloitte-uk-plateauing-at-the-peak-the-state-of-the-smartphone.pdf, 2019.

- [34] Android Developers. Permissions updates in Android 11 | Android Developers. <https://developer.android.com/about/versions/11/privacy/permissions>. Last Accessed: October 28, 2022.
- [35] Android Developers. Request app permissions | Android Developers. <https://developer.android.com/training/permissions/requesting>. Last Accessed: October 28, 2022.
- [36] Grace Dillon. An In-Depth Look at In-Game Mobile Advertising. *ExchangeWire*, August 2021. <https://www.exchangewire.com/deep-dive/an-in-depth-look-at-in-game-mobile-advertising>.
- [37] Disconnect, Inc. Disconnect. <https://disconnect.me/disconnect>. Last Accessed: October 28, 2022.
- [38] Claire Dolin, Ben Weinshel, Shawn Shan, Chang Min Hahn, Euirim Choi, Michelle L. Mazurek, and Blase Ur. Unpacking Perceptions of Data-Driven Inferences Underlying Online Targeting and Personalization. In *ACM Conference on Human Factors in Computing Systems*, CHI '18, pages 493:1–493:12, Montreal, Quebec, Canada, April 2018. ACM.
- [39] Julia Earp and Jessica Staddon. “I Had No Idea This Was a Thing”: On the Importance of Understanding the User Experience of Personalized Transparency Tools. In *Workshop on Socio-Technical Aspects in Security and Trust*, STAST '16, pages 79–86, Los Angeles, California, USA, December 2016. ACM.
- [40] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. Ask the Experts: What Should Be on an IoT Privacy and Security Label? In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 447–464, San Jose, CA, USA, May 2020. IEEE. ISSN: 2375-1207.
- [41] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. Which Privacy and Security Attributes Most Impact Consumers’ Risk Perception and Willingness to Purchase IoT Devices? In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 519–536, San Francisco, CA, USA, May 2021. IEEE.
- [42] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, page 1388–1401, New York, NY, USA, 2016. Association for Computing Machinery.
- [43] Facebook, Inc. Facebook Q4 2020 Earnings – Slides. <https://investor.fb.com/investor-events/event-details/2021/Facebook-Q4-2020-Earnings-default.aspx>, January 2021.

- [44] Florian M. Farke, David G. Balash, Maximilian Golla, Markus Dürmuth, and Adam J. Aviv. Are privacy dashboards good for end users? evaluating user perceptions and reactions to google’s my activity. In *USENIX Security Symposium*, 2021.
- [45] Shehroze Farooqi, Maaz Musa, Zubair Shafiq, and Fareed Zaffar. CanaryTrap: Detecting Data Misuse by Third-Party Apps on Online Social Networks. *Proceedings on Privacy Enhancing Technologies*, 2020.
- [46] Adrienne Felt and David Evans. Privacy Protection for Social Networking APIs. In *The Web 2.0 Security and Privacy*, 2008.
- [47] Adrienne Porter Felt, Erika Chin, Steve Hanna, Dawn Song, and David Wagner. Android Permissions Demystified. In *ACM Conference on Computer and Communications Security*, 2011.
- [48] Adrienne Porter Felt, Erika Chin, Steve Hanna, Dawn Song, and David Wagner. Android permissions demystified. In *Proceedings of the 18th ACM Conference on Computer and Communications Security*, CCS ’11, page 627–638, New York, NY, USA, 2011. Association for Computing Machinery.
- [49] Adrienne Porter Felt, Serge Egelman, Matthew Finifter, Devdatta Akhawe, and David Wagner. How to Ask for Permission. In *USENIX Conference on Hot Topics in Security*, 2012.
- [50] Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. Android permissions: User attention, comprehension, and behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, SOUPS ’12, New York, NY, USA, 2012. Association for Computing Machinery.
- [51] Simone Fischer-Hübner, Julio Angulo, Farzaneh Karegar, and Tobias Pulls. Transparency, Privacy and Trust – Technology for Tracking and Controlling My Data Disclosures: Does This Work? In *IFIP International Conference on Trust Management*, IFIP TM ’16, pages 3–14, Darmstadt, Germany, July 2016. IFIP.
- [52] Jen Fitzpatrick. Putting You in Control of Your Data. <https://blog.google/technology/safety-security/our-work-keep-you-safe>, May 2021.
- [53] FDA Center for Devices and Radiological Health. Device Labeling. <https://www.fda.gov/medical-devices/overview-device-regulation/device-labeling>, October 2020.
- [54] FDA Center for Food Safety and Applied Nutrition. How to Understand and Use the Nutrition Facts Label. <https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label>, February 2022.
- [55] Geoffrey A. Fowler. I checked Apple’s new privacy ‘nutrition labels.’ Many were false. *The Washington Post*, January 2021. <https://www.washingtonpost.com/technology/2021/01/29/apple-privacy-nutrition-label/>.

- [56] Thomas Franke. P3P - platform for privacy preferences project. *Wirtschaftsinf.*, 43(2):197–199, 2001.
- [57] Mohammad Ghasemisharif, Amrutha Ramesh, Stephen Checkoway, Chris Kanich, and Jason Polakis. O Single Sign-Off, Where Art Thou? An Empirical Analysis of Single Sign-On Account Hijacking and Session Management on the Web. In *USENIX Security Symposium*, 2018.
- [58] Ghostery. Ghostery, best Ad Blocker & Privacy Browser. <https://www.ghostery.com>. Last Accessed: October 28, 2022.
- [59] Joshua Gluck, Florian Schaub, Amy Friedman, Hana Habib, Norman Sadeh, Lorie Faith Cranor, and Yuvraj Agarwal. How Short Is Too Short? Implications of Length and Framing on the Effectiveness of Privacy Notices. In *Symposium on Usable Privacy and Security*, SOUPS '16, pages 321–340, Denver, Colorado, USA, July 2016. USENIX.
- [60] Google. How Google helps you share data safely with third-party apps & services. <https://support.google.com/accounts/answer/10130420>. Last Accessed: May 10, 2021.
- [61] Google. Manage third-party apps & services with access to your account. <https://support.google.com/accounts/answer/3466521>. Last Accessed: May 10, 2021.
- [62] Google. More granular Google Account permissions with Google OAuth and APIs. <https://developers.googleblog.com/2018/10/more-granular-google-account.html>. Last Accessed: September 20, 2021.
- [63] Google. Provide information for Google Play’s Data safety section - Play Console Help. <https://support.google.com/googleplay/android-developer/answer/10787469>. Last Accessed: October 28, 2022.
- [64] Google. Use your Google Account to sign in to other apps or services. <https://support.google.com/accounts/answer/112802>. Last Accessed: May 10, 2021.
- [65] Google. Google – Ad Settings. <https://adssettings.google.com>, March 2009. Last Accessed: October 28, 2022.
- [66] Google. Google – My Activity. <https://myactivity.google.com>, June 2016. Last Accessed: October 28, 2022.
- [67] Thomas Groß. Validity and Reliability of the Scale Internet Users’ Information Privacy Concerns (IUIPC). *Proceedings on Privacy Enhancing Technologies*, 2021.
- [68] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorie Faith Cranor, Norman Sadeh, and Florian Schaub. “It’s a Scavenger Hunt”: Usability of Websites’ Opt-Out and Data Deletion Choices. In *ACM Conference*

on Human Factors in Computing Systems, CHI '20, pages 1–12, Honolulu, Hawaii, USA, April 2020. ACM.

- [69] Hana Habib, Yixin Zou, Aditi Jannu, Neha Sridhar, Chelse Swoopes, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. An Empirical Analysis of Data Deletion and Opt-out Choices on 150 Websites. In *Symposium on Usable Privacy and Security*, SOUPS '19, pages 387–406, Santa Clara, California, USA, August 2019. USENIX.
- [70] Catherine Han, Irwin Reyes, Amit Elazari Bar On, Joel Reardon, Álvaro Feal, Serge Egelman, Narseo Vallina-Rodriguez, et al. Do you get what you pay for? comparing the privacy behaviors of free vs. paid apps. In *Workshop on Technology and Consumer Protection (ConPro 2019), in conjunction with the 39th IEEE Symposium on Security and Privacy, 23 May 2019, San Francisco, CA, USA.*, San Francisco, CA, USA, 2019. IEEE.
- [71] Catherine Han, Irwin Reyes, Álvaro Feal, Joel Reardon, Primal Wijesekera, Narseo Vallina-Rodriguez, Amit Elazari Bar On, Kenneth A. Bamberger, and Serge Egelman. The price is (not) right: Comparing privacy in free and paid apps. *Proc. Priv. Enhancing Technol.*, 2020(3):222–242, 2020.
- [72] Julia Hanson, Miranda Wei, Sophie Veys, Matthew Kugler, Lior Strahilevitz, and Blase Ur. Taking Data Out of Context to Hyper-Personalize Ads: Crowdworkers' Privacy Perceptions and Decisions to Disclose Private Information. In *ACM Conference on Human Factors in Computing Systems*, CHI '20, pages 1–13, Honolulu, Hawaii, USA, April 2020. ACM.
- [73] Eelco Herder and Olaf van Maaren. Privacy Dashboards: The Impact of the Type of Personal Data and User Control on Trust and Perceived Risk. In *ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '20, pages 169–174, Virtual Conference, July 2020. ACM.
- [74] Pili Hu, Ronghai Yang, Yue Li, and Wing Cheong Lau. Application Impersonation: Problems of OAuth and API Design in Online Social Networks. In *ACM Conference on Online Social Networks*, 2014.
- [75] Markus Huber, Martin Mulazzani, Sebastian Schrittwieser, and Edgar Weippl. Appinspect: Large-Scale Evaluation of Social Networking Apps. In *ACM Conference on Online Social Networks*, 2013.
- [76] Umar Iqbal, Steven Englehardt, and Zubair Shafiq. Fingerprinting the Fingerprinters: Learning to Detect Browser Fingerprinting Behaviors. In *IEEE Symposium on Security and Privacy*, SP '21, pages 283–301, Virtual Conference, May 2021. IEEE.
- [77] Jim Isaak and Mina J. Hanna. User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer*, 2018.

- [78] Milena Janic, Jan Pieter Wijnnga, and Thijs Veugen. Transparency Enhancing Tools (TETs): An Overview. In *Workshop on Socio-Technical Aspects in Security and Trust, STAST '13*, pages 18–25, New Orleans, Louisiana, USA, June 2013. IEEE.
- [79] Carlos Jensen and Colin Potts. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 471–478, 2004.
- [80] Garrett A. Johnson, Scott K. Shriver, and Shaoyin Du. Consumer Privacy Choice in Online Advertising: Who Opt Out and at What Cost to Industry? *Marketing Science*, 39(1):33–51, January 2020.
- [81] Jacob Kastrenakes. Apple says there are now over 1 billion active iPhones. *The Verge*, January 2022. <https://www.theverge.com/2021/1/27/22253162/iphone-users-total-number-billion-apple-tim-cook-q1-2021>.
- [82] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A “Nutrition Label” for Privacy. In *Symposium on Usable Privacy and Security, SOUPS '09*, pages 4:1–4:12, Mountain View, California, USA, July 2009. ACM.
- [83] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security, SOUPS '09*, pages 1–12, New York, NY, USA, July 2009. Association for Computing Machinery.
- [84] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. Standardizing privacy notices: an online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10*, pages 1573–1582, New York, NY, USA, April 2010. Association for Computing Machinery.
- [85] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3393–3402, Paris France, April 2013. ACM.
- [86] Mohammad Taha Khan, Joe DeBlasio, Geoffrey M. Voelker, Alex C. Snoeren, Chris Kanich, and Narseo Vallina-Rodriguez. An empirical analysis of the commercial VPN ecosystem. In *Proceedings of the Internet Measurement Conference 2018, IMC 2018, Boston, MA, USA, October 31 - November 02, 2018*, pages 443–456, Boston, MA, USA, 2018. ACM.
- [87] Konrad Kollnig, Pierre Dewitte, Max Van Kleek, Ge Wang, Daniel Omeiza, Helena Webb, and Nigel Shadbolt. A fait accompli? an empirical study into the absence of consent to Third-Party tracking in android apps. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 181–196, Virtual Conference, August 2021. USENIX Association.
- [88] Konrad Kollnig, Anastasia Shuba, Reuben Binns, Max Van Kleek, and Nigel Shadbolt. Are iPhones really better for privacy? a comparative study of iOS and Android apps. *Proceedings on Privacy Enhancing Technologies*, 2022(2):6–4, 2022.

- [89] Konrad Kollnig, Anastasia Shuba, Max Van Kleek, Reuben Binns, and Nigel Shadbolt. Goodbye Tracking? Impact of iOS App Tracking Transparency and Privacy Labels. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, Virtual Conference, April 2022. Association for Computing Machinery. arXiv: 2204.03556.
- [90] Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor. Why Johnny Can't Opt out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising. In *ACM Conference on Human Factors in Computing Systems*, CHI '12, pages 589–598, Austin, Texas, USA, May 2012. ACM.
- [91] Pedro Giovanni Leon, Blase Ur, Yang Wang, Manya Sleeper, Rebecca Balebako, Richard Shay, Lujo Bauer, Mihai Christodorescu, and Lorrie Faith Cranor. What Matters to Users? Factors That Affect Users' Willingness to Share Information with Online Advertisers. In *Symposium on Usable Privacy and Security*, SOUPS '13, pages 7:1–7:12, Newcastle, United Kingdom, July 2013. ACM.
- [92] Adam Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In *25th USENIX Security Symposium (USENIX Security 16)*, Austin, TX, August 2016. USENIX Association.
- [93] Tianshi Li, Kayla Reiman, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. Understanding challenges for developers to create accurate privacy nutrition labels. In *CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.
- [94] Yucheng Li, Deyuan Chen, Tianshi Li, Yuvraj Agarwal, Lorrie Faith Cranor, and Jason I. Hong. Understanding ios privacy nutrition labels: An exploratory large-scale analysis of app store data. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, CHI EA '22, New York, NY, USA, 2022. Association for Computing Machinery.
- [95] Delfina Malandrino, Vittorio Scarano, and Raffaele Spinelli. How Increased Awareness Can Impact Attitudes and Behaviors toward Online Privacy Protection. In *IEEE Conference on Social Computing*, SocialCom '13, pages 57–62, Alexandria, Virginia, USA, September 2013. IEEE.
- [96] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, 15(4):336–355, December 2004.
- [97] Célestin Matte, Nataliia Bielova, and Cristiana Santos. Do Cookie Banners Respect my Choice?: Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework. In *IEEE Symposium on Security and Privacy*, SP '20, pages 791–809, Virtual Conference, May 2020. IEEE.

- [98] Aleecia M. McDonald and Lorrie Faith Cranor. The Cost of Reading Privacy Policies. *HeinOnline*, 4(3):543–568, 2009.
- [99] Maryam Mehrnezhad. A cross-platform evaluation of privacy notices and tracking practices. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 97–106, Genoa, Italy, 2020. IEEE.
- [100] Meta. Facebook Login. <https://developers.facebook.com/docs/facebook-login>. Last Accessed: October 28, 2022.
- [101] Chance Miller. Apple Says App Tracking Transparency Feature Will Launch in ‘Early Spring’ with iOS 14 Update, January 2021. <https://9to5mac.com/2021/01/27/app-tracking-transparency-spring>.
- [102] Eric Miraglia. Privacy That Works for Everyone, May 2019. <https://www.blog.google/technology/safety-security/privacy-everyone-io>.
- [103] Patrick Murmann and Simone Fischer-Hübner. Tools for Achieving Usable Ex Post Transparency: A Survey. *IEEE Access*, 5:22965–22991, October 2017.
- [104] Lily Hay Newman. Google Will Delete Your Data by Default–In 18 Months. *Wired*, June 2020. <https://www.wired.com/story/google-auto-delete-data>.
- [105] Casey Newton. Jumbo: Is a Powerful Privacy Assistant for iOS That Cleans up Your Social Profiles. *The Verge*, April 2019. <https://www.theverge.com/2019/4/9/18300775>.
- [106] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence. In *ACM Conference on Human Factors in Computing Systems, CHI ’20*, pages 1–13, Honolulu, Hawaii, USA, April 2020. ACM.
- [107] Ofce of the California Attorney General. California Consumer Privacy Act (CCPA): Final Text of Proposed Regulations., 2020. <https://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/oal-sub-final-text-of-regs.pdf>.
- [108] Nathan Olivarez-Giles. How to Use Google’s New My Activity Privacy Tool. *The Wall Street Journal*, July 2016. <https://www.wsj.com/articles/how-to-use-googles-new-my-activity-privacy-tool-1467402973>.
- [109] Nikolaos Pantelaios, Nick Nikiforakis, and Alexandros Kapravelos. You’ve changed: Detecting malicious browser extensions through their update deltas. In Jay Ligatti, Xinming Ou, Jonathan Katz, and Giovanni Vigna, editors, *CCS ’20: 2020 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, USA, November 9-13, 2020*, pages 477–491, Virtual Conference, 2020. ACM.
- [110] European Parliament. Regulation (EU) 2016/679 of the European Parliament and of the Council, 2016. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>.

- [111] Nicole Lyn Pesce. Americans trust amazon and google more than the police or the government. *Marketwatch*, Jan. 18, 2020. <https://www.marketwatch.com/story/people-trust-amazon-and-google-more-than-the-police-or-the-government-2020-01-14>.
- [112] David Pierce. Google’s New Magic Number for Storing Personal Data: 18 Months. *Protocol*, June 2020. <https://www.protocol.com/google-delete-data-18-months>.
- [113] Robert Pitofsky. Prepared Statement of the Federal Trade Commission On Self Regulation and Privacy Online. Technical Report 106th Congress, Federal Trade Commission, Senate Commerce Committee, July 1999.
- [114] Robert Pitofsky. Privacy Online: Fair Information Practices in the Electronic Marketplace: A Federal Trade Commission Report to Congress. Technical report, Federal Trade Commission, May 2000.
- [115] Emilee Rader. Awareness of Behavioral Tracking and Information Privacy Concern in Facebook and Google. In *Symposium on Usable Privacy and Security*, SOUPS ’14, pages 51–67, Menlo Park, California, USA, July 2014. USENIX.
- [116] Emilee Rader, Samantha Hautea, and Anjali Munasinghe. “I Have a Narrow Thought Process”: Constraints on Explanations Connecting Inferences and Self-Perceptions. In *Symposium on Usable Privacy and Security*, SOUPS ’20, pages 457–488, Virtual Conference, August 2020. USENIX.
- [117] Ashwini Rao, Florian Schaub, and Norman Sadeh. What do they know about me? Contents and Concerns of Online Behavioral Profiles. In *ASE International Conference on Privacy, Security, Risk and Trust*, PASSAT ’14, pages 1–13, Cambridge, Massachusetts, USA, December 2014. ASE.
- [118] Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, and Ruogu Kang. Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online. In *Symposium on Usable Privacy and Security*, SOUPS ’16, pages 77–96, Denver, Colorado, USA, July 2016. USENIX.
- [119] Philip Raschke, Axel Küpper, Olha Drozd, and Sabrina Kirrane. Designing a GDPR-Compliant and Usable Privacy Dashboard. In *IFIP International Summer School on Privacy and Identity Management*, IFIP SC ’17, pages 221–236, Ispra, Italy, September 2017. IFIP.
- [120] Joel Reardon, Álvaro Feal, Primal Wijesekera, Amit Elazari Bar On, Narseo Vallina-Rodriguez, and Serge Egelman. 50 ways to leak your data: An exploration of apps’ circumvention of the android permissions system. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 603–620, Santa Clara, CA, August 2019. USENIX Association.

- [121] Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek. How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples. In *IEEE Symposium on Security and Privacy*, SP '19, pages 1326–1343, San Francisco, California, USA, May 2019. IEEE.
- [122] Joel R Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T Graves, Fei Liu, Aleecia McDonald, Thomas B Norton, and Rohan Ramanath. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Tech. LJ*, 30:39, 2015.
- [123] Jingjing Ren, Martina Lindorfer, Daniel J Dubois, Ashwin Rao, David Choffnes, and Narseo Vallina-Rodriguez. Bug Fixes, Improvements, ... and Privacy Leaks – A Longitudinal Study of PII Leaks Across Android App Versions. In *Proceedings of the 25th Network and Distributed System Security Symposium (NDSS 2018)*, San Diego, CA, USA, 2018. Internet Society.
- [124] Irwin Reyes and Michael Lack. Research Talk – API Privacy: A Look at G Suite Marketplace Permissions and Policies. *Workshop on Technology and Consumer Protection*, 2020.
- [125] N. Cameron Russell, Florian Schaub, Allison McDonald, and William Sierra-Rocafort. APIs and Your Privacy. *Fordham University Center on Law and Information Policy CLIP*, 2019.
- [126] Aafaq Sabir, Evan Lafontaine, and Anupam Das. Analyzing the Impact and Accuracy of Facebook Activity on Facebook's Ad-Interest Inference Process. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):76:1–76:34, April 2022.
- [127] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A Design Space for Effective Privacy Notices. In *Symposium on Usable Privacy and Security*, SOUPS '15, pages 1–17, Ottawa, Canada, July 2015. USENIX.
- [128] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 1–17, Ottawa, July 2015. USENIX Association.
- [129] Florian Schaub, Aditya Marella, Pranshu Kalvani, Blase Ur, Chao Pan, Emily Forney, and Lorrie Faith Cranor. Watching Them Watching Me: Browser Extensions' Impact on User Privacy Awareness and Concern. In *Workshop on Usable Security*, USEC '16, San Diego, California, USA, February 2016. ISOC.
- [130] Florian Schaub, Aditya Marella, Pranshu Kalvani, Blase Ur, Chao Pan, Emily Forney, and Lorrie Faith Cranor. Watching them watching me: Browser extensions' impact on user privacy awareness and concern. In *NDSS workshop on usable security*, pages 1–10, 2016.
- [131] Allison Schiff. The Gaming Audience Is A Lot More Diverse And Desirable Than You Think, May 2020.

- [132] Sebastian Schnorf, Martin Ortlieb, and Nikhil Sharma. Trust, Transparency & Control in Inferred User Interest Models. In *ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 2449–2454, Toronto, Ontario, Canada, April 2014. ACM.
- [133] Mike Schroepfer. An Update on Our Plans to Restrict Data Access on Facebook. *Facebook Newsroom*, Apr 2018. <https://about.fb.com/news/2018/04/restricting-data-access/>.
- [134] Gian Luca Scoccia, Marco Autili, Giovanni Stilo, and Paola Inverardi. An empirical study of privacy labels on the Apple iOS mobile app store. In *IEEE/ACM 9th International Conference on Mobile Software Engineering and Systems (MOBILESoft '22)*, page 11, Pittsburgh, PA, USA, 2022. ACM.
- [135] Suranga Seneviratne, Harini Kolamunna, and Aruna Seneviratne. A measurement study of tracking in paid mobile applications. In *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, New York, NY, USA, June 22-26, 2015, pages 7:1–7:6, New York, NY, USA, 2015. ACM.
- [136] Yun Shen, Pierre-Antoine Vervier, and Gianluca Stringhini. Understanding worldwide private information collection on android. In *28th Annual Network and Distributed System Security Symposium, NDSS 2021, virtually, February 21-25, 2021*, Virtual Conference, 2021. The Internet Society.
- [137] Daniel Smullen, Yuanyuan Feng, Shikun Aerin Zhang, and Norman Sadeh. The Best of Both Worlds: Mitigating Trade-offs Between Accuracy and User Burden in Capturing Mobile App Privacy Preferences. *Proceedings on Privacy Enhancing Technologies*, 2020, 2020.
- [138] FTC Staff. Protecting consumer privacy in an era of rapid change—a proposed framework for businesses and policymakers. *Journal of Privacy and Confidentiality*, 3(1), 2011.
- [139] San-Tsai Sun and Konstantin Beznosov. The Devil is in the (Implementation) Details: An Empirical Analysis of OAuth SSO Systems. In *ACM Conference on Computer and Communications Security*, 2012.
- [140] Herman T. Tavani. Philosophical Theories of Privacy: Implications for an Adequate Online Privacy Policy. *Metaphilosophy*, 38(1):1–22, January 2007.
- [141] Michael Carl Tschantz, Serge Egelman, Jaeyoung Choi, Nicholas Weaver, and Gerald Friedland. The Accuracy of the Demographic Inferences Shown on Google’s Ad Settings. In *Workshop on Privacy in the Electronic Society*, WPES '18, pages 33–41, Toronto, Canada, October 2018. ACM.
- [142] Twitter. Log in with Twitter | Docs | Twitter Developer Platform. <https://developer.twitter.com/en/docs/authentication/guides/log-in-with-twitter>. Last Accessed: October 28, 2022.

- [143] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. In *Symposium on Usable Privacy and Security*, SOUPS '12, pages 4:1–4:15, Washington, D.C., USA, July 2012. ACM.
- [144] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. (Un)informed Consent: Studying GDPR Consent Notices in the Field. In *ACM Conference on Computer and Communications Security*, CCS '19, pages 973–990, London, United Kingdom, November 2019. ACM.
- [145] Antoine Vastel, Pierre Laperdrix, Walter Rudametkin, and Romain Rouvoy. Fp-Scanner: The Privacy Implications of Browser Fingerprint Inconsistencies. In *USENIX Security Symposium*, SSYM '18, pages 135–150, Baltimore, Maryland, USA, August 2018. USENIX.
- [146] Sophie Veys, Daniel Serrano, Madison Stamos, Margot Herman, Nathan Reitinger, Michelle L. Mazurek, and Blase Ur. Pursuing Usable and Useful Data Downloads Under GDPR/CCPA Access Rights via Co-Design. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, pages 217–242, August 2021.
- [147] Haoyu Wang, Zhe Liu, Jingyue Liang, Narseo Vallina-Rodriguez, Yao Guo, Li Li, Juan Tapiador, Jincun Cao, and Guoai Xu. Beyond google play: A large-scale comparative study of chinese android app markets. In *Proceedings of the Internet Measurement Conference 2018, IMC 2018, Boston, MA, USA, October 31 - November 02, 2018*, pages 293–307, Boston, MA, USA, 2018. ACM.
- [148] Huiyi Wang, Liu Wang, and Haoyu Wang. Market-level analysis of government-backed covid-19 contact tracing apps. In *2020 35th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*, pages 79–84, Virtual Conference, 2020. Association for Computing Machinery.
- [149] Na Wang, Heng Xu, and Jens Grossklags. Third-Party Apps on Facebook: Privacy and the Illusion of Control. In *ACM Symposium on Computer Human Interaction for Management of Information Technology*, 2011.
- [150] Jeffrey Warshaw, Nina Taft, and Allison Woodruff. Intuitions, Analytics, and Killing Ants: Inference Literacy of High School-educated Adults in the {US}. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, pages 271–285, 2016.
- [151] Miranda Wei, Madison Stamos, Sophie Veys, Nathan Reitinger, Justin Goodman, Margot Herman, Dorota Filipczuk, Ben Weinshel, Michelle L. Mazurek, and Blase Ur. What Twitter Knows: Characterizing Ad Targeting Practices, User Perceptions, and Ad Explanations Through Users' Own Twitter Data. In *USENIX Security Symposium*, SSYM '20, pages 145–162, Virtual Conference, August 2020. USENIX.
- [152] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L. Mazurek, and Blase Ur. Oh, the Places You've Been! User Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing.

In *ACM Conference on Computer and Communications Security*, CCS '19, pages 149–166, London, United Kingdom, November 2019. ACM.

- [153] Primal Wijesekera, Arjun Baokar, Ashkan Hosseini, Serge Egelman, David Wagner, and Konstantin Beznosov. Android Permissions Remystified: A Field Study on Contextual Integrity. In *USENIX Security Symposium*), 2015.
- [154] Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David Wagner, and Konstantin Beznosov. The Feasibility of Dynamically Granted Permissions: Aligning Mobile Privacy with User Preferences. In *IEEE Symposium on Security and Privacy*, 2017.
- [155] Craig E. Wills and Can Tatar. Understanding what they do with what they know. In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, WPES '12, pages 13–18, New York, NY, USA, October 2012. Association for Computing Machinery.
- [156] Yuxi Wu, Panya Gupta, Miranda Wei, Yasemin Acar, Sascha Fahl, and Blase Ur. Your Secrets Are Safe: How Browsers' Explanations Impact Misconceptions About Private Browsing Mode. In *The World Wide Web Conference*, WWW '18, pages 217–226, Lyon, France, April 2018. ACM.
- [157] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. How usable are ios app privacy labels? *Proceedings on Privacy Enhancing Technologies*, 4:204–228, 2022.
- [158] Yuchen Zhou and David Evans. SSOScan: Automated Testing of Web Applications for Single Sign-On Vulnerabilities. In *USENIX Security Symposium*, 2014.
- [159] Christian Zimmermann, Rafael Accorsi, and Günter Müller. Privacy Dashboards: Reconciling Data-Driven Business Models and Privacy. In *IEEE Conference on Availability, Reliability and Security*, ARES '14, pages 152–157, Fribourg, Switzerland, September 2014. IEEE.
- [160] Shoshana Zuboff. *Age of surveillance capitalism: the fight for a human future at the new frontier of power*. *The age of surveillance capitalism : the fight for a human future at the new frontier of power /*. PublicAffairs, New York, first edition, 2019.
- [161] Chaoshun Zuo, Qingchuan Zhao, and Zhiqiang Lin. AuthScope: Towards Automatic Discovery of Vulnerable Authorizations in Online Services. In *ACM Conference on Computer and Communications Security*, 2017.
- [162] Álvaro Feal, Paolo Calciati, Narseo Vallina-Rodriguez, Carmela Troncoso, and Alessandra Gorla. Angel or devil? a privacy study of mobile parental control apps. *Proceedings on Privacy Enhancing Technologies*, 2020(2):314–335, 2020.

Appendix A: Google My Activity

A.1 Screening Survey Instrument

Thank you for your interest in our survey. Your answers are important to us. **Please read the following instructions carefully:** (i) Take your time in reading and answering the questions. (ii) Answer the questions as accurately as possible. (iii) It is okay to say that you don't know an answer.

- S1** Do you have a personal Gmail address (an email address ending in "gmail.com")?
☐ Yes ☐ No

- S2** How long do you have that Gmail address?
☐ Less than a year ☐ More than five years
☐ One year ☐ I do not have a Gmail address
☐ Three years ☐ Unsure
☐ Five years

- S3** Which other Google products do you currently use? (Select all that apply.)
- | | | |
|-------------------------------------|-------------------------------------|--------------------------------------|
| <input type="radio"/> Gmail | <input type="radio"/> Google Search | <input type="radio"/> Google Pay |
| <input type="radio"/> Google Maps | <input type="radio"/> Google Play | <input type="radio"/> Android device |
| <input type="radio"/> YouTube | <input type="radio"/> Google Drive | <input type="radio"/> None of these |
| <input type="radio"/> Google Chrome | <input type="radio"/> Google News | |

- A0** Google began in January 1996 as a research project. Its initial public offering took place on August 19, 2004. Did the initial public offering of Google take place in 1996?
☐ Yes ☐ No
☐ Other (please specify) _____

- S4** How frequently do you use these products? *[Included only products selected in S3.]*

	Always	Often	Sometimes	Rarely	Never	Unsure
Gmail	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Maps	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
YouTube	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Chrome	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Search	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Play	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Drive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google News	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Pay	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Android device	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

- S5** How important is using Google products to your Internet experience?
☐ Not important ☐ Important
☐ Slightly important ☐ Very important
☐ Moderately important

[Next, we asked the 10 × UIIPC items of the control, awareness, and collection scale as described by Malhotra et al. [96].]

A.1.1 Main Survey Instrument

- Q1** How aware are you of the amount of information that Google is collecting about your activities online?
☐ Not at all aware ☐ Moderately aware
☐ Slightly aware ☐ Extremely aware
☐ Somewhat aware

- Q2** How concerned are you with the amount of information Google is collecting about your activities online?

- ☐ Not at all concerned
☐ Slightly concerned
☐ Somewhat concerned

- ☐ Moderately concerned
☐ Extremely concerned

Q2_A Please explain why.

Answer: _____

Q3 How often do you benefit from the amount of information that Google collects about your activities online?

- ☐ Never
☐ Rarely
☐ Sometimes

 ☐ Often
☐ Always

Q3_A Please explain why.

Answer: _____

Some people use strategies to limit the amount of information that companies can collect about them online.

Q4 Do you have any strategies for managing the kind of information Google may collect about you?

Answer: _____

What is Google My Activity?

The following briefly introduces you to Google's My Activity page. For every account, Google provides an overview called My Activity, which contains the history of activities of your interactions with Google products. Below is Google's description of the My Activity page.

- What is My Activity?
 - *"My Activity is a central place to view and manage activity like searches you've done, websites you've visited, and videos you've watched."*¹
- What are activities?
 - *"When you use certain Google services, like Search, YouTube, or Chrome, your activity can be saved as data to your account."*¹

Google Login Page

This survey requires that you login to your primary Google account for accessing items in your My Activity page.

Privacy Note: We do not track or store your email address as part of this study, and we will not be able to tie your email address to any results or analysis. The researchers will never see your email address. At no time do the researchers have access to your Google account.

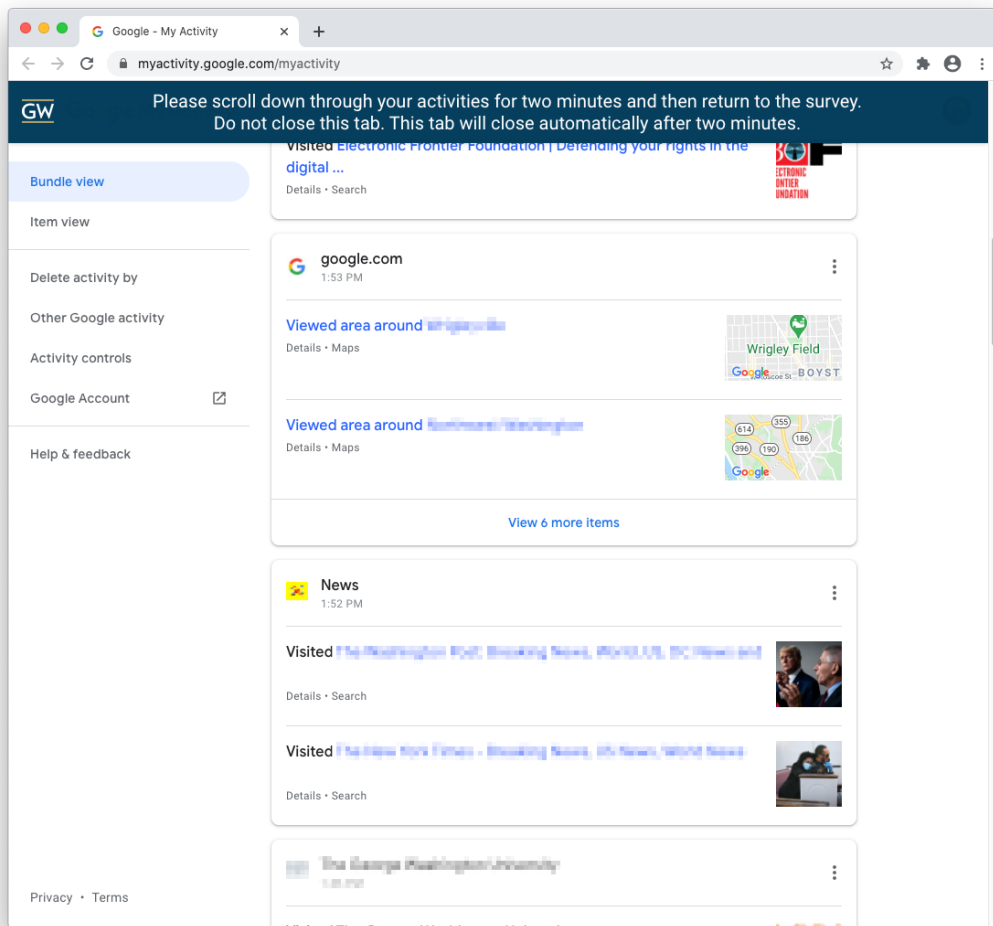
Explore My Activity

In the next part of the study, we will ask you to explore Google's My Activity page for your Google account. You will have an opportunity to interact with your Google My Activity page for one minute and will then be returned to the survey.

Privacy Note: We do not track or store your email address as part of this study, and we will not be able to tie your email address to any results or analysis. The researchers will never see your email address. At no time do the researchers have access to your Google account.

Participants explored their My Activity page.

¹<https://support.google.com/accounts/answer/7028918>, as of October 19, 2022



Q5 Please provide any immediate reactions you have to exploring the My Activity page.

Answer: _____

Q6 Have you visited the My Activity page prior to this study?

☐ Yes ☐ No ☐ Unsure

A1 What is the shape of a red ball?

☐ Red ☐ Blue
☐ Round ☐ Square

Q7 Provide three purposes for which you think Google is using your activity data.

1. _____ 2. _____ 3. _____

Q7_A Based on your answer before, do you believe the purposes for Google using this information is beneficial to you in any way?

	Not at all	Slightly	Somewhat	Moderately	Extremely
<i>Beneficial</i>					
1. Purpose	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Purpose	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Purpose	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q7_B Based on your answer before, do you believe the purposes for Google using this information is harmful to you in any way?

	Not at all	Slightly	Somewhat	Moderately	Extremely
<i>Harmful</i>					
1. Purpose	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Purpose	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Purpose	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q8 Are there any other concerns you might have with Google collecting this information?

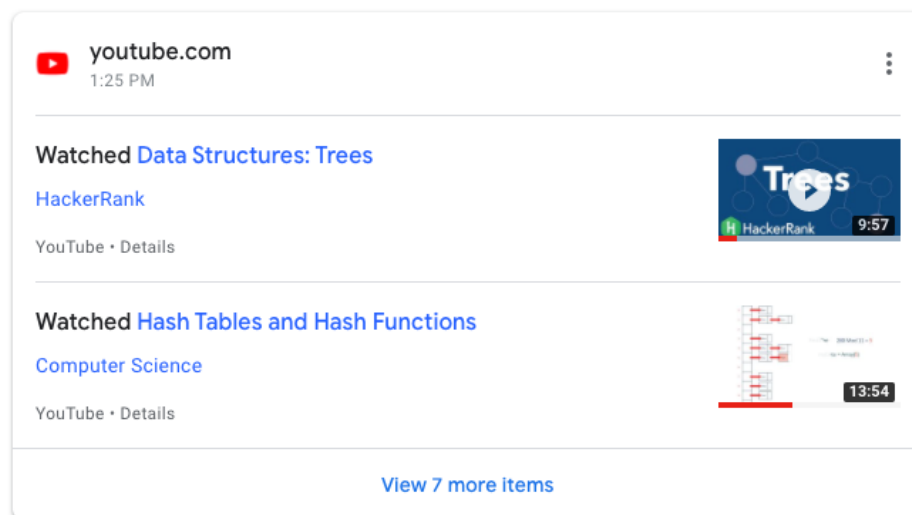
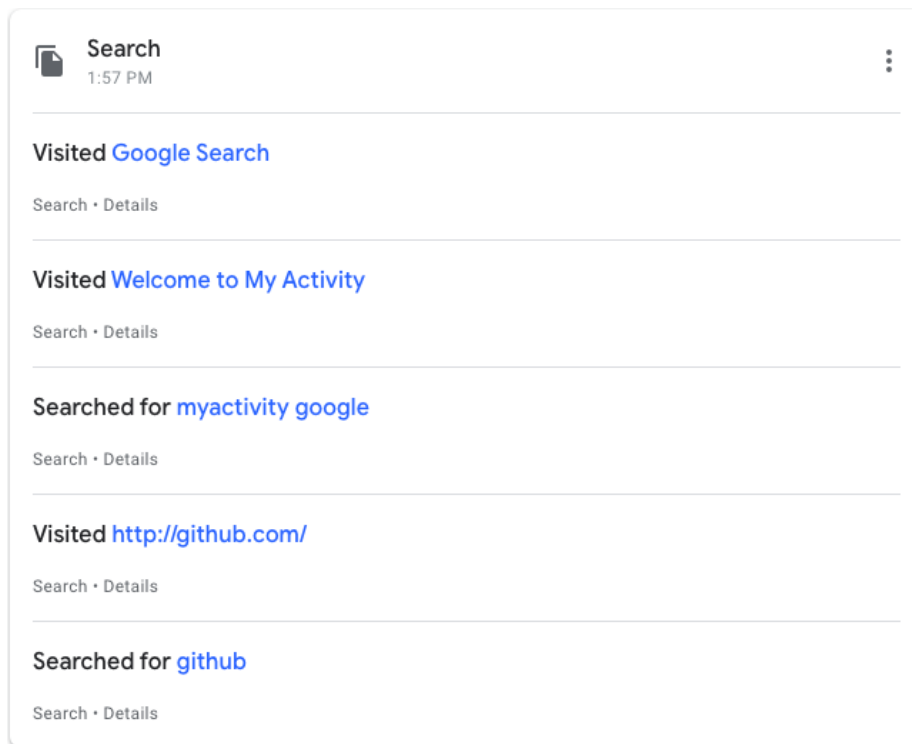
Answer: _____

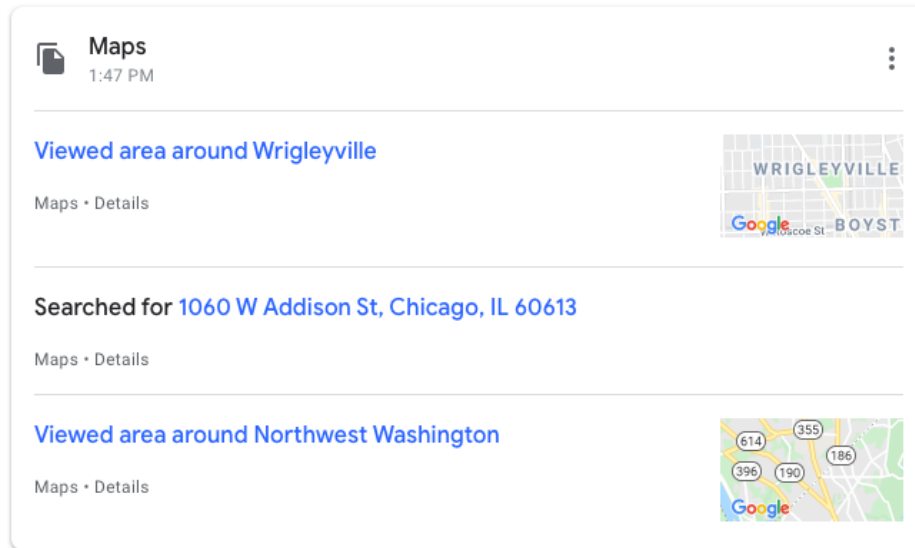
Q9 Do you believe your experience using Google services is improved by Google collecting this information?

- | | |
|--|--------------------------------------|
| <input type="radio"/> Strongly disagree | <input type="radio"/> Agree |
| <input type="radio"/> Disagree | <input type="radio"/> Strongly agree |
| <input type="radio"/> Neither agree nor disagree | |

Activity Presentation

In the next part of the survey, we will ask you questions about **nine activities** from your My Activity page. The activities are chosen randomly. We do not collect information about that activity as part of this survey. That information remains private, **only accessible to you** and Google. We only note which service the activity is associated with, e. g., “Google search” vs. “YouTube view”, and the date on which it occurred. Further details are not collected as part of this survey.





Q10 Do you recall this activity?

- ☐ Yes ☐ No

Q11 Prior to seeing this activity, have you been aware that Google stored this activity?

- ☐ Yes ☐ No ☐ Unsure

Q12 Storing this activity is necessary for my experience with using *[Google product name]*.

- ☐ Strongly disagree ☐ Agree
☐ Disagree ☐ Strongly agree
☐ Neither agree nor disagree

Q13 Storing this activity changes my experience with using *[Google product name]* in the following way:

- ☐ Greatly harms my experience
☐ Harms my experience
☐ Slightly harms my experience
☐ Does not change my experience
☐ Slightly improves my experience
☐ Improves my experience
☐ Greatly improves my experience

Q14 If you were to change how long this data is stored, when would you want it to be deleted?

- ☐ Immediately, I do not want this data to be collected
☐ After a few hours
☐ After a day
☐ After a week
☐ After a month
☐ After 3 months
☐ After 18 months
☐ I wouldn't delete

[Repeat questions Q10 to Q14 for each activity presented to the participant]

Q15 Describe two feelings you had after viewing the activities we showed you.

1. _____ 2. _____

Q16 If Google offered a paid plan where they do not collect your activity data but you received the same features and user experience from their products, how much would you be willing to pay for such a service?

[Slider from \$0 to \$100 per month]

Activity Explanations

Google gives different explanations for why they collect activity data. They differentiate between three categories of activities: Web activities, YouTube activities, and Maps activities. The following shows for each of these categories the explanations Google gives for why they store activity data.

- Q17** Do you think this is an appropriate reason to store your Google Search activity?
- | | |
|--|--|
| <input type="radio"/> Absolutely inappropriate | <input type="radio"/> Slightly appropriate |
| <input type="radio"/> Inappropriate | <input type="radio"/> Appropriate |
| <input type="radio"/> Slightly inappropriate | <input type="radio"/> Absolutely appropriate |
| <input type="radio"/> Neutral | |
- Q18** Do you think this is an appropriate reason to store your YouTube activity?
- | | |
|--|--|
| <input type="radio"/> Absolutely inappropriate | <input type="radio"/> Slightly appropriate |
| <input type="radio"/> Inappropriate | <input type="radio"/> Appropriate |
| <input type="radio"/> Slightly inappropriate | <input type="radio"/> Absolutely appropriate |
| <input type="radio"/> Neutral | |
- Q19** Do you think this is an appropriate reason to store your Google Maps activity?
- | | |
|--|--|
| <input type="radio"/> Absolutely inappropriate | <input type="radio"/> Slightly appropriate |
| <input type="radio"/> Inappropriate | <input type="radio"/> Appropriate |
| <input type="radio"/> Slightly inappropriate | <input type="radio"/> Absolutely appropriate |
| <input type="radio"/> Neutral | |

Auto-Delete Options

Google allows you to change how long your online activity is stored. These settings are called auto-delete options and can be used to automatically delete activities older than a set amount of time.

Q20 Would you like to change how long your activities are stored?

Q20_A Please explain if and why you would like to change how long your activities are stored?

Answer: _____

Q21 Google provides a way for you to pause collection of activity data, what do you believe happens when you pause activity data collection?

- ☐ Google no longer collects activity data about me
☐ Google still collects activity data about me, but does not associate it with my account
☐ Google still collects activity data about me and still associates it with my account, but simply does not display it on the My Activity page.
☐ Other: _____

A2 What is the color of a red ball?

- | | |
|-----------------------------|------------------------------|
| <input type="radio"/> Red | <input type="radio"/> Blue |
| <input type="radio"/> Round | <input type="radio"/> Square |

Q22 Do you think My Activity helps you to better understand what data Google collects about you?

- | | |
|--|--------------------------------------|
| <input type="radio"/> Strongly disagree | <input type="radio"/> Agree |
| <input type="radio"/> Disagree | <input type="radio"/> Strongly agree |
| <input type="radio"/> Neither agree nor disagree | |

Q22_A Please explain why.

Answer: _____

Q23 After completing this survey, do you see yourself changing any setting in your My Activity page?

- | | | |
|---------------------------|--------------------------|------------------------------|
| <input type="radio"/> Yes | <input type="radio"/> No | <input type="radio"/> Unsure |
|---------------------------|--------------------------|------------------------------|

Q23_A Which setting, if any, would you change?

Answer: _____

Q24 In a month, do you see yourself reviewing and/or deleting activities using My Activity?

- | | | |
|---------------------------|--------------------------|------------------------------|
| <input type="radio"/> Yes | <input type="radio"/> No | <input type="radio"/> Unsure |
|---------------------------|--------------------------|------------------------------|

Q24_A Which kinds of activities, if any, would you review and/or delete?

Answer: _____

Q25 Now that you have explored, My Activity, do you plan using Google products differently in the future?

☐ Yes

☐ No

☐ Unsure

Q25_A What would you change and why?

[Shown only if answer to Q25 was "Yes"]

Answer: _____

Q25_B Why are you unsure if you would change using Google products?

[Shown only if answer to Q25 was "Unsure"]

Answer: _____

Q25_C Why would you not change using Google products?

[Shown only if answer to Q25 was "No"]

Answer: _____

Q26 How concerned are you with the amount of information Google is collecting about your activities online?

☐ Not at all concerned

☐ Moderately concerned

☐ Slightly concerned

☐ Extremely concerned

☐ Somewhat concerned

Q26_A Please explain why.

Answer: _____

Q27 How often do you benefit from the amount of information that Google collects about your activities online?

☐ Never

☐ Often

☐ Rarely

☐ Always

☐ Sometimes

Q27_A Please explain why.

Answer: _____

D1 What is your gender?

☐ Woman

☐ Prefer not to disclose

☐ Man

☐ Prefer to self-describe

☐ Non-binary

D2 What is your age?

☐ 18 – 24

☐ 55 – 64

☐ 25 – 34

☐ 65 or older

☐ 35 – 44

☐ Prefer not to disclose

☐ 45 – 54

D3 What is the highest degree or level of school you have completed?

☐ No schooling completed

☐ Some high school, no diploma

☐ High school graduate, diploma, or equivalent

☐ Some college credit, no degree

☐ Trade / technical / vocational training

☐ Associate degree

☐ Bachelor's degree

☐ Master's degree

☐ Professional degree (e. g., J.D., M.D.)

☐ Doctorate degree

☐ Prefer not to disclose

D4 Which of the following best describes your educational background or job field?

☐ I have an education in, or work in, the field of computer science, computer engineering or IT.

☐ I do not have an education in, nor do I work in, the field of computer science, computer engineering or IT.

☐ Prefer not to disclose

A.2 Qualitative Codes

- **ad-blocker (1)**
- **against-collection (25)**
activities, computer-usage, maps, not-necessary, phone-usage, risky, search-history, selling-data, website-history, youtube
- **against-data-collection (5)**
- **against-filter-bubble (2)**
- **against-targeted-advertising (14)**
- **amount (85)**
as-expected, less-than-expected, more-than-expected
- **artificial-intelligence-concerns (3)**
biased-algorithms
- **behavior-modification (1)**
- **being-used-to-make-money (5)**
- **benefit (152)**
- **better-understanding (115)**
google-profit-motive, how-long-information-is-stored, how-much-information-is-collected, how-to-change-settings, implications-of-data-collection, inferences, myactivity-existence, other-google-apps, usage-history, what-google-is-doing, what-information-is-collected, when-information-is-collected, why-information-is-collected
- **browser (38)**
ad-blocker, block-cookies, clear-history-cache, extension, popup-blocker, private-window, remove-cookies, second-browser, tracker-blocking
- **change-settings (98)**
all-settings, auto-delete, automaticaly-delete-data, data-storage-length, delete-activities-more-often, delete-frequency, google-my-activity, history, how-long-information-stored, limit-data-collection, maps, never-delete, reminders, search, stop-data-collection, turn-off-collection, turn-on-more-data-collection, youtube
- **change-use (16)**
more-careful, use-non-google-browser, use-non-google-search, use-non-google-services
- **collection-aware (11)**

- **collection-beneficial (67)**
better-recommendations, easy-to-find-activity, improved-experience, improves-services, maps, personalization, personalized-ads, preferences, recommendations, revisit-activities, search, search-history, shortcuts, youtube
- **collection-not-beneficial (5)**
- **collection-unnecessary (2)**
- **concern (6)**
- **control (1)**
- **convenient (22)**
but-not-enough
- **cookies (1)**
- **creepy (4)**
- **data-collection (93)**
interests, internet-use-history, location, user-information
- **data-collection-beneficial (1)**
- **delete-activity (90)**
all, app-data, history, irrelevant, maps, music, search, sensitive, web, youtube
- **delete-data (12)**
all, banking, history, web-activity, work-related-information
- **delete-immediately (5)**
web-history
- **delete-information (3)**
google
- **details (17)**
more-than-expected
- **did-not-change-perspective (5)**
- **dont-store-activities-time-period (40)**
day-or-two, forever, long, maps, one-month, one-week, short-period, three-months
- **experience-improvements (109)**
ease-of-use, maps, personalization, saving-history, search, user-interface

- **false-inferences (4)**
- **google-monopoly (4)**
- **google-products-convenient (5)**
google-account-login
- **google-products-necessary (18)**
better-than-alternatives, daily, habit, habits, life, school, tied-to-google, work
- **government (1)**
- **happy-with-google (3)**
- **happy-with-status-quo (37)**
- **harm (1)**
- **history (43)**
useful
- **i-am-the-product (1)**
- **increase-google-product-use (2)**
- **increased-awareness (3)**
after-exploring-myactivity, google-data-collection, search
- **interested (9)**
- **liability-coverage (1)**
- **limit-information (25)**
amount, gdpr-opt-out, location, sensitive, sharing
- **limit-usage (21)**
account, accounts, anonymize-activity, gmail, google-products, google-chrome, google-products, google-services, no-chrome, no-click-ads, no-store-information, restrict-search, search, third-party-aps
- **little-benefit (26)**
- **make-money (68)**
sell-data
- **makes-note-of-offending-website (1)**
- **modify-settings-or-configuration (17)**
ad-personalization-off, data-gathering, google, privacy-settings

- **my-activity-helpful (16)**
- **myactivity-useful (14)**
activity-history, detailed-information, for-google, metrics, one-location-for-activity-history, recommendations, search
- **never-delete (2)**
maps, youtube
- **no-answer (1)**
- **no-benefit (45)**
- **no-change (105)**
already-configured, likes-current-settings, too-time-consuming, would-forget
- **no-sentiment (40)**
- **no-strategies (78)**
doesnt-know-how, overwhelmed, wants-strategies
- **no-surprise (34)**
- **none (1)**
- **not-better-understanding (13)**
already-know-activity
- **not-relevant-to-store (42)**
a-few-months, eighteen-months, five-years, maps, one-month, one-week, one-year, search, three-months, youtube
- **nothing-to-hide (23)**
- **personalized-ads (84)**
creepy, dislike, manipulation, not-useful, partial-useful, useful
- **physical-safety-concerns (2)**
location-data
- **privacy-aware (36)**
amount-of-data-collected

- **privacy-concerns (253)**

amount-of-information, behavior-modification, biographical-information, control-data, data-collection, data-future, data-sensitive, false-inferences, feels-violated, google-pervasive, government, how-long-data-stored, inferences, information-collection, information-used-against-me, information-used-against-you, invasion-of-privacy, invasive, lack-of-consent, lack-of-control, location, monitoring, not-properly-anonomized-data, search-history-used-against-me, selling-data, selling-information, sensitive-information, stalking, third-parties, tracking, uncomfortable-sharing

- **privacy-protection (1)**

- **privacy-resigned (34)**

change-wont-make-difference, data-collection-unavoidable

- **privacy-tradeoff (19)**

convenience, free-services

- **product-improvements (42)**

auto-complete, products-and-services, search

- **profit-form-data (14)**

get-paid, little-in-return, services-are-free

- **protect-google (1)**

- **request-data (1)**

- **research (15)**

machine-learning

- **review (11)**

google-data-collection-policies, google-myactivity, history, work-related

- **scared (2)**

- **search-bubble (3)**

- **security (9)**

accounts, passwords

- **security-concerns (85)**

data-breach, data-future, data-misuse, data-released, location

- **skeptical (31)**

designed-to-confuse, does-not-show-all-data-collection, google-only-helping-itself, undisclosed-use-of-data

- **social-influence (1)**
- **societal-impact (2)**
- **some-benefit (55)**
- **strategies (1)**
fake-info
- **suggestions (149)**
not-useful, partial-useful, useful
- **suggestions-not-accurate (1)**
maps, youtube
- **surprise (51)**
- **targeted-advertising (123)**
- **too-much-data (2)**
- **too-much-time-required-to-change (5)**
- **tools (17)**
antimalware, google-activity-control, google-privacy-checkup, non-tracking-search-engine, secure-server, vpn
- **trust-google (33)**
- **uncertain (12)**
about-benefit, how-much-is-collected, what-is-collected
- **unclear (8)**
- **unclear-statement (15)**
- **uncomfortable (4)**
- **unconcerned (131)**
collection, have-control, my-activity-not-a-risk, non-sensitive
- **undecided (57)**
how-much-work-to-change, if-necessary, needs-more-review, privacy-level, wants-to-review-data-collection, wants-to-review-devices, wants-to-review-history, wants-to-review-settings, what-information-to-save, what-settings-to-change
- **undisclosed-collection (3)**
- **unknown (7)**

- **unknowns (98)**

amount-data-collected, how-information-collected, how-information-used, how-long-information-stored, how-much-information, how-often-data-collected, how-secure-is-information, how-to-control, how-to-protect-privacy, inferences, information-collected, information-future, risk-versus-benefit, what-data-collected, what-information-collected, when-information-collected, where-information-collected, where-information-goes, who-has-access, why-information-collected, why-information-stored, why-myactivity, why-recent-web-search-not-shown

- **unsure (1)**

- **use-false-information (6)**

alias, fake-account

- **wants-control-over-what-is-stored (4)**

A.3 Demographics

Table A.1: Full demographic data of the participants of the main survey.

		Male		Female		Other		Total	
		n	%	n	%	n	%	n	%
Age	18–24	14	9	13	8	2	1	29	19
	25–34	20	13	15	10	0	0	35	23
	35–44	17	11	14	9	0	0	31	20
	45–54	14	9	15	10	0	0	29	19
	55–64	12	8	12	8	0	0	24	16
	65 or older	2	1	2	1	0	0	4	3
	No answer	0	0	0	0	1	1	1	1
Education	Some High School	1	1	0	0	0	0	1	1
	High School	12	8	2	1	1	1	15	10
	Some College	15	10	15	10	1	1	31	20
	Training	1	1	1	1	0	0	2	1
	Associates	7	5	8	5	0	0	15	10
	Bachelor’s	28	18	27	18	0	0	55	36
	Master’s	11	7	13	8	0	0	24	16
	Professional	1	1	1	1	0	0	2	1
	Doctorate	3	2	4	3	0	0	7	5
	No answer	0	0	0	0	1	1	1	1
Back-ground	Tech	29	19	8	5	0	0	37	24
	No Tech	48	31	59	39	2	1	109	71
	No answer	2	1	4	3	1	1	7	5
Total		79	52	71	46	3	2	153	100

A.4 Screenshots of My Activity and the Survey

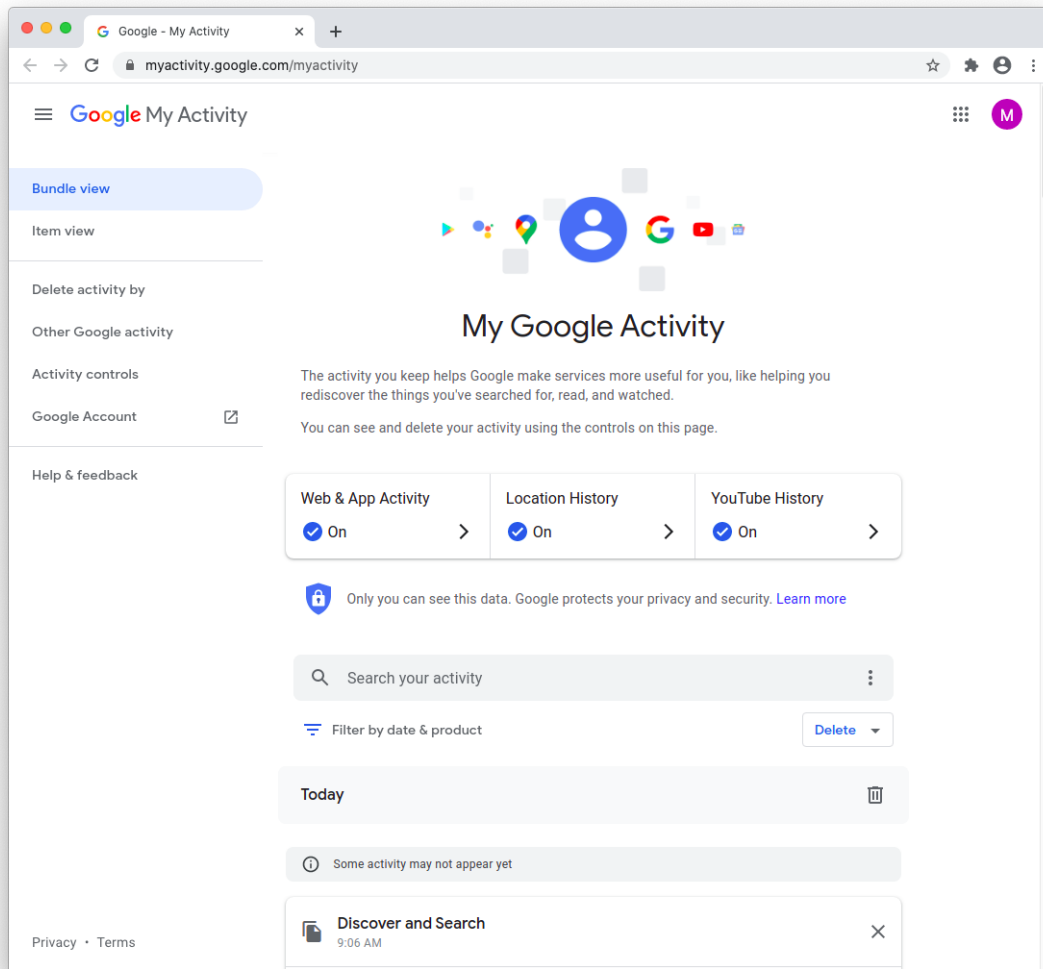


Figure A.1: The My Activity user interface at the time of the study in September 2020.

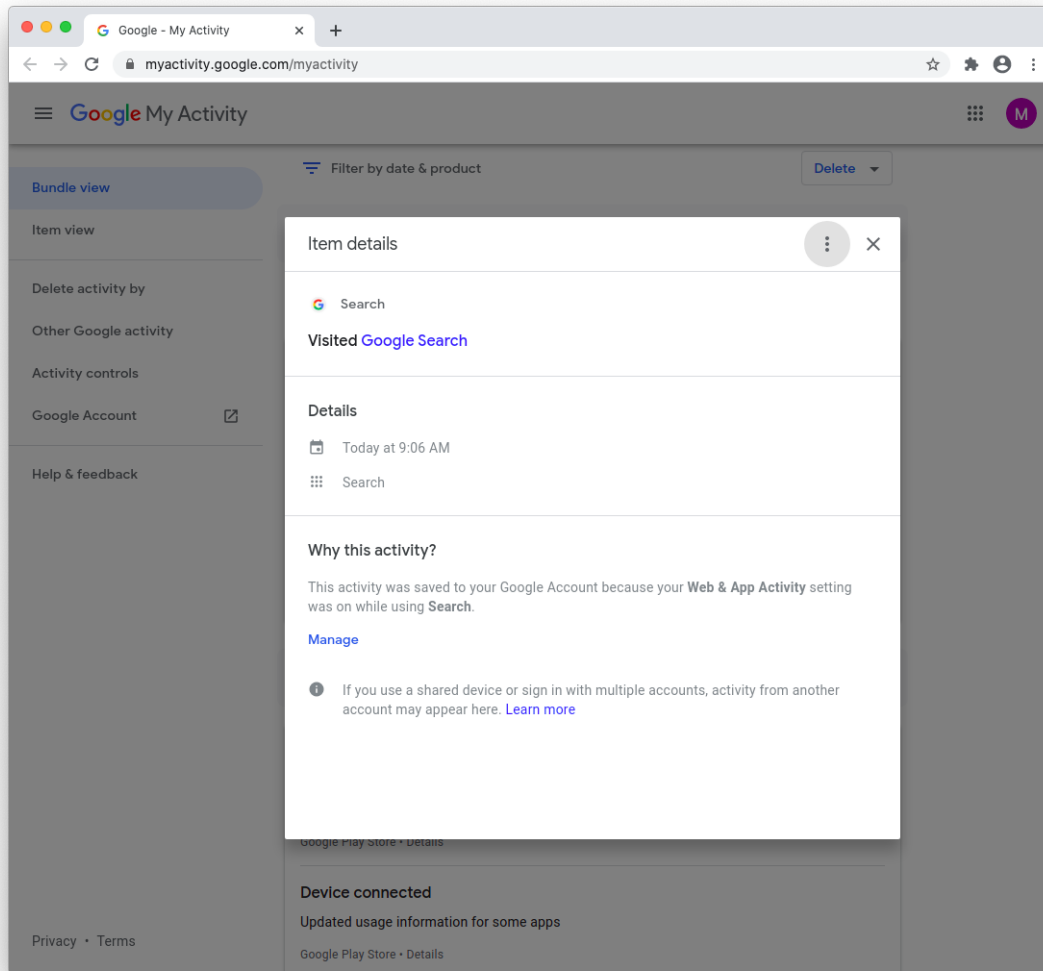


Figure A.2: Clicking on the details link of an activity opens an additional view providing information like the time and date, type of the activity, and in which service or app the activity was collected.

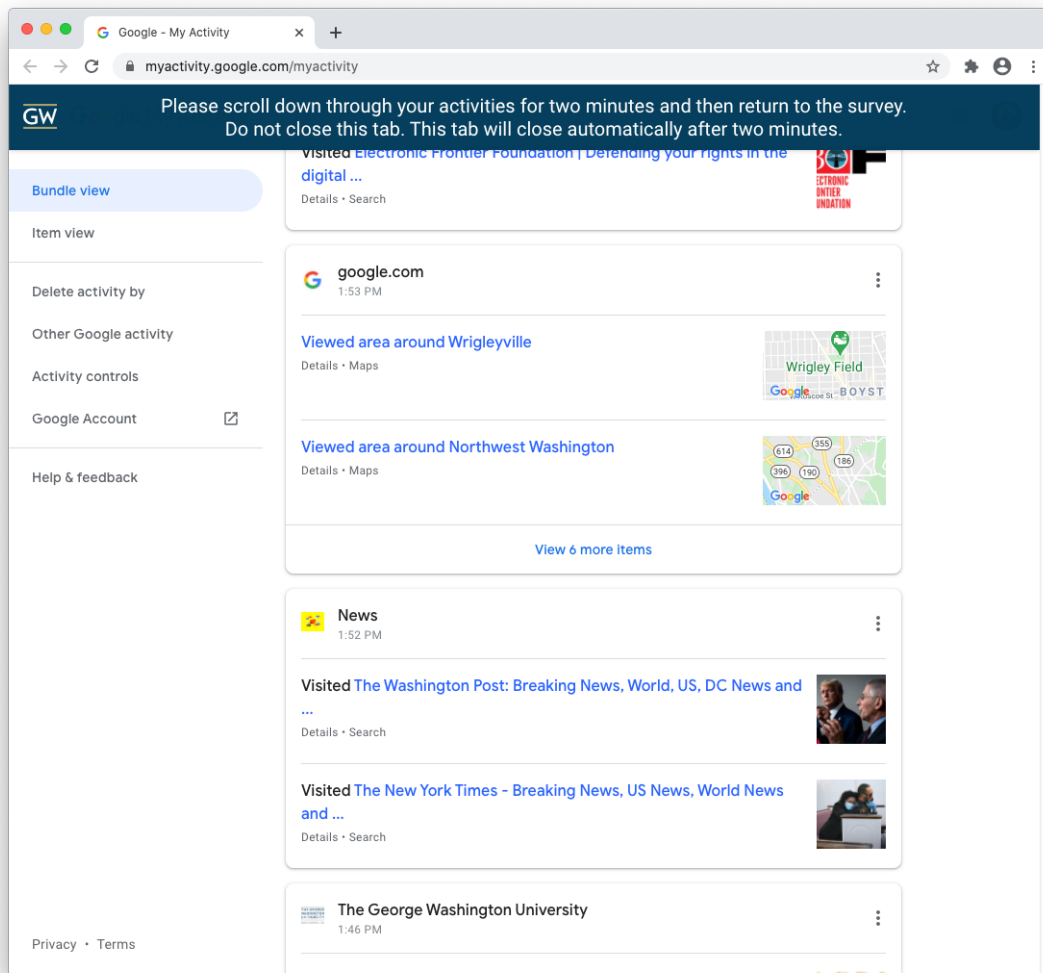


Figure A.3: During the My Activity exploration phase of the survey, we added a banner at the top of the My Activity page and disabled all buttons and hyperlinks on the page.

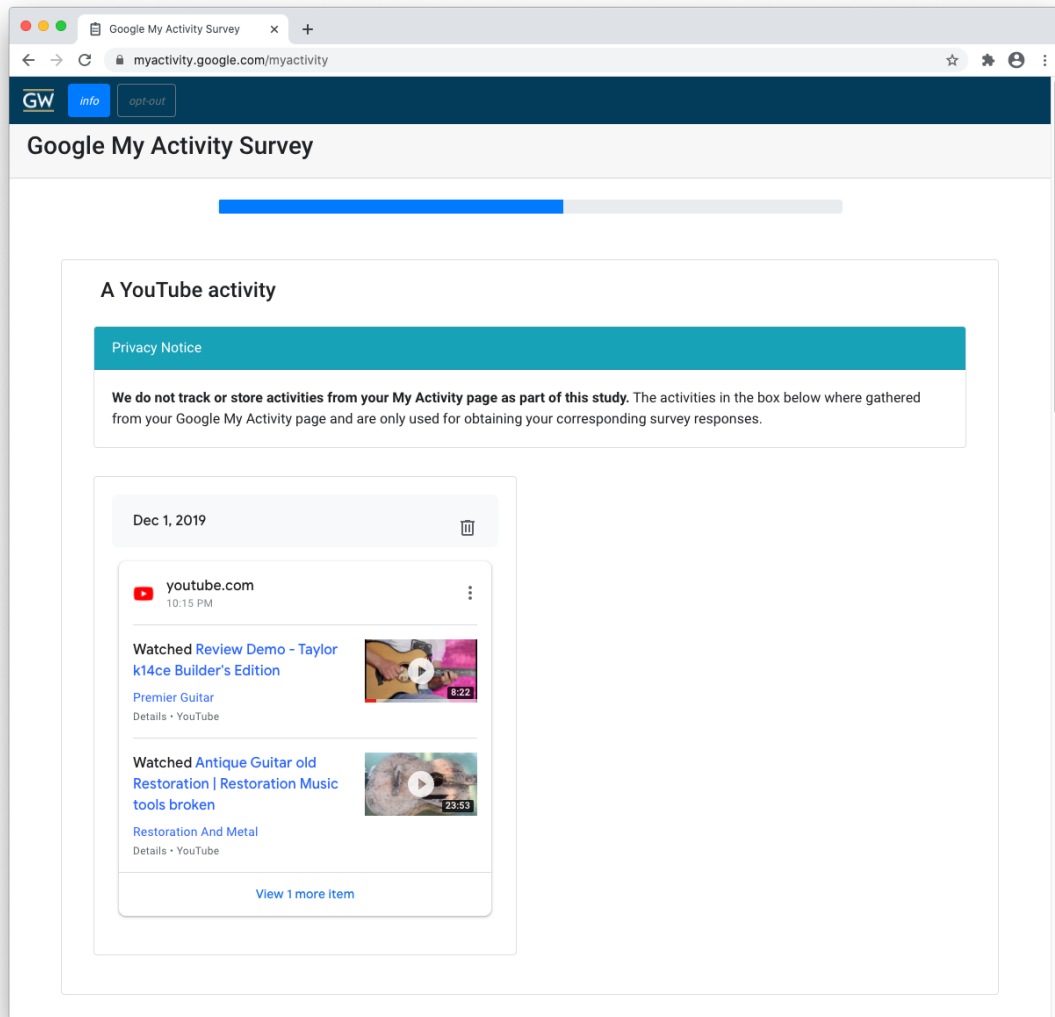


Figure A.4: In the Activity Presentation phase of the survey, we embedded activity bundles into the survey page. The participants saw this type of page up to 9 times. Every time with another activity bundle.

Appendix B: API Privacy

B.1 Survey Instrument

B.1.1 First Survey

Please read the following instructions carefully:

- Take your time in reading and answering the questions.
- Answer the questions as accurately as possible.

Q₁1 Do you have a Gmail (Google) account?

- ☐ Yes ☐ No

Privacy Notice: We do not transmit your email address to our server as part of this study, and we will not be able to tie your email address to any results or analysis. All uses of your email address are local to your browser. The researchers will never see your email address. At no time do the researchers have access to your Google account.

[Q₁2 through Q₁13 are shown if Q₁1 is “Yes”]

Q₁2 Do you use {*participant email address*} as your primary Gmail (Google) account?

- ☐ Yes
☐ No, I use a different Gmail (Google) as my primary account

Q₁3 Whose Gmail address is it?

- ☐ It is my own account. I have sole ownership of this account.
☐ It is an institutional account. A business, school, or organization gave it to me.
☐ It is my shared account. I share the account with someone else (e.g., a partner or family member).
☐ It is someone else’s account. Someone else has sole ownership of this account.

Q₁4 How long have you had this Gmail address as your primary Gmail (Google) account?

- ☐ Less than 1 month ☐ Less than or about 10 years
☐ Less than or about 1 year ☐ More than 10 years
☐ Less than or about 5 years ☐ Unsure

Q₁5 What is the color of a red ball?

- ☐ Red ☐ Round ☐ Blue ☐ Square

Use your Google Account to sign in to other apps or services

- You can use your Google Account to sign in to third-party apps and services.
- You won’t have to remember individual usernames and passwords for each account.
- Third-party apps and services are created by companies or developers that aren’t Google.

Q₁₆ Do you recall ever using your Google Account to sign in to third-party apps or services as described above?

☐ Yes

☐ No

☐ Unsure

Q₁₇ Thinking about the last time you used your Google Account to sign into a third-party app or service, what app or service did you use your Google Account to sign into? *[Shown if Q₁₆ is “Yes”]* (short answer)

Q₁₈ Thinking about the last time you used your Google Account to sign into a third-party app or service, what did you consider before signing in using your Google Account? *[Shown if Q₁₆ is “Yes”]* (short answer)

Q₁₉ If you were given the option to use your Google account to sign into a third-party app or service, what would you consider before using this feature? *[Shown if Q₁₆ is “No” or “Unsure”]* (short answer)

Manage third-party apps & services with access to your Google Account

- Google lets you give third-party apps and services access to different parts of your Google Account.
- Third-party apps and services are created by companies or developers that aren’t Google.
- For example, you may download an app that helps you schedule workouts with friends. This app may request access to your Google Calendar and Contacts to suggest times and friends for you to meet up with.

Q₁₁₀ Do you recall ever granting a third-party app access to your Google Account as described above?

☐ Yes

☐ No

☐ Unsure

Q₁₁₁ Thinking about the last time you granted a third-party app access to your Google Account, what was the purpose of allowing that access? *[Shown if Q₁₁₀ is “Yes”]* (short answer)

Q₁₁₂ Thinking about the last time you granted a third-party app access to your Google Account, what did you consider before granting a third-party app access to your Google account? *[Shown if Q₁₁₀ is “Yes”]* (short answer)

Q₁₁₃ If you were given the option to grant a third-party app access to your Google Account, what would you consider before granting access? *[Shown if Q₁₁₀ is “No” or “Unsure”]* (short answer)

[Q₁₁₄ through Q₁₂₃ are shown if Q₁₁ is “No”]

Use an existing online account to sign in to other apps or services

- You can use your existing account on various online platforms (e.g., Facebook, Google, Apple, etc.) to sign in to third-party apps and services.
- You won’t have to remember individual usernames and passwords for each account.

- Third-party apps and services are created by companies or developers that aren't part of those platforms.

Q₁14 Do you recall ever using an existing online account to sign in to a third-party app or service as described above?

☐ Yes ☐ No ☐ Unsure

Q₁15 Thinking about the last time you used an existing online account to sign into a third-party app or service, what app or service did you sign into? [*Shown if Q₁14 is "Yes"*] (short answer)

Q₁16 Thinking about the last time you used an existing online account to sign into a third-party app or service, what online account did you use? [*Shown if Q₁14 is "Yes"*] (short answer)

Q₁17 Thinking about the last time you used an existing online account to sign into a third-party app or service, what did you consider before signing in using that existing online account? [*Shown if Q₁14 is "Yes"*] (short answer)

Q₁18 If you were given the option to use an existing online account to sign into a third-party app or service, what online account would you use? [*Shown if Q₁14 is "No" or "Unsure"*] (short answer)

Q₁19 If you were given the option to use an existing online account to sign into a third-party app or service, what would you consider before using this feature? [*Shown if Q₁14 is "No" or "Unsure"*] (short answer)

Manage third-party apps & services with access to your online Account

- Online platforms let you give third-party apps and services access to different parts of your accounts on those platforms.
- Third-party apps and services are created by companies or developers that aren't part of those platforms.
- For example, you may download an app that helps you schedule workouts with friends. This app may request access to your online calendar (e.g., Google Calendar, Apple iCloud Calendar, Microsoft Outlook, etc.) and contact list to suggest times and friends for you to meet up with.

Q₁20 Do you recall ever granting a third-party app access to one of your online accounts as described above?

☐ Yes ☐ No ☐ Unsure

Q₁21 Thinking about the last time you granted a third-party app access to one of your online accounts, what was the purpose of allowing that access? [*Shown if Q₁20 is "Yes"*] (short answer)

Q₁22 Thinking about the last time you granted a third-party app access to one of your online accounts, what did you consider before granting a third-party app access to one of your online accounts? [*Shown if Q₁20 is "Yes"*] (short answer)

Q123 If you were given the option to grant a third-party app access to one of your online accounts, what would you consider before granting access? [Shown if **Q120** is “No” or “Unsure”] (short answer)

These questions were followed by the 8 IUIPC items as described by Malhotra et al. [96] and Groß [67].

D1 What is your gender?

- | | |
|----------------------------------|---|
| <input type="radio"/> Woman | <input type="radio"/> Prefer not to disclose |
| <input type="radio"/> Man | <input type="radio"/> Prefer to self-describe |
| <input type="radio"/> Non-binary | |

D2 What is your age?

- | | | | |
|-------------------------------|-------------------------------|-----------------------------------|--|
| <input type="radio"/> 18 – 24 | <input type="radio"/> 35 – 44 | <input type="radio"/> 55 – 64 | <input type="radio"/> Prefer not to disclose |
| <input type="radio"/> 25 – 34 | <input type="radio"/> 45 – 54 | <input type="radio"/> 65 or older | |

D3 What is the highest degree or level of school you have completed?

- ☐ No schooling completed
- ☐ Some high school
- ☐ High school
- ☐ Some college credit, no degree
- ☐ Trade, technical, or vocational training
- ☐ Associate degree
- ☐ Bachelor’s degree
- ☐ Master’s degree
- ☐ Professional degree
- ☐ Doctorate degree
- ☐ Prefer not to disclose

D4 Which of the following best describes your educational background or job field?

- ☐ I have an education in, or work in, the field of computer science, computer engineering or IT.
- ☐ I do not have an education in, nor do I work in, the field of computer science, computer engineering or IT.
- ☐ Prefer not to disclose

B.1.2 Second Survey

About Account Linking

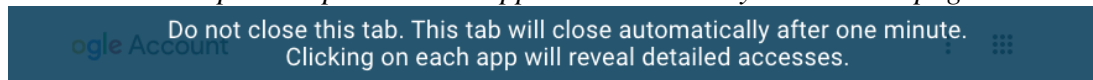
- Users of online services can often link their accounts to products made by other companies, which can operate on behalf of the user’s account. For example, a user can link their Spotify account (i.e., a music streaming service) to their Amazon Alexa (i.e., a voice-controlled smart speaker) in order to play their favorite music.
- Your Gmail account, as a Google account, can also be linked in this way. In the following questions, we ask about how your Google account can similarly be linked to

other (non-Google) services. Later we will use a web browser extension to do further analysis of how you may (or may not) link your Google account.

Explore Apps With Access To Your Account

- In the next part of the study, we will ask you to explore Google’s “Apps with access to your account” page for your Google account.
- You will have an opportunity to interact with your Google’s “Apps with access to your account” page for **one minute** and will then be returned to the survey.
- We have **disabled clicking** for various links and buttons on the Google’s “Apps with access to your account” page to help you focus better on the access that third party applications have to your Google account.




Participants explored their Apps with access to your account page.



← Apps with access to your account

Third-party apps with account access

You gave these sites and apps access to some of your Google Account data, including info that may be sensitive. Remove access for those you no longer trust or use. [Learn about the risks](#)

 Creately	Has access to Google Drive
 Mozilla Thunderbird Email	Has access to Gmail
 Zoom	Has access to Google Calendar




Signing in with Google

You use your Google Account to sign in to these sites and apps. They can view your name, email address, and profile picture. [Learn more](#)

Google Account sign-in prompts

Allow Google to offer a faster way to sign in with your Google Account on supported third-party sites

☒

 Creately
 Chicago Tribune
 The New York Times


Use your Google Account to sign in to other apps or services


- You can use your Google Account to sign in to third-party apps and services. You won't have to remember individual usernames and passwords for each account.
- Third-party apps and services are created by companies or developers that aren't Google.


Participants view a list of SSOs authorized to their Google account.


Signing in with Google

You use your Google Account to sign in to these sites and apps. They can view your name, email address, and profile picture.

 **Creately**

 **Dropbox for Gmail**

 **Chicago Tribune**

 **The New York Times**

Q21 You have [*n*] third-party apps and services that you can **sign into using your Google account**. Before using my Google account to sign into a website or third-party app, I consider:

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
...how secure the app or website is.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...how the app or website will use your data.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...whether you can delete your data from the app or website.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...whether the app or website will tell you if something changes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...who else can see your data on the app or website.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q22 How often do you review what services you can sign into using your Google account?

- | | |
|------------------------------|-------------------------------|
| <input type="radio"/> Never | <input type="radio"/> Weekly |
| <input type="radio"/> Rarely | <input type="radio"/> Monthly |
| <input type="radio"/> Daily | <input type="radio"/> Yearly |

Manage third-party apps & services with access to your Google Account





- Google lets you give third-party apps and services access to different parts of your Google Account.
- Third-party apps and services are created by companies or developers that aren't Google.

- For example, you may download an app that helps you schedule workouts with friends. This app may request access to your Google Calendar and Contacts to suggest times and friends for you to meet up with.

Participants view third-party apps with access to their Google account.

Third-party apps with account access

You gave these sites and apps access to some of your Google Account data, including info that may be sensitive. Remove access for those you no longer trust or use.

 Creately	Has access to Google Drive
 Dropbox for Gmail	Has access to Gmail, Google Contacts
 Mozilla Thunderbird Email	Has access to Gmail
 Zoom	Has access to Google Calendar

Q₂₃ You have [n] third-party apps and services that **have access to your Google account data**. Before granting a website or third-party app access to my Google account, I consider:

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
...how secure the app or website is.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...how the app or website will use your data.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...whether you can delete your data from the app or website.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...whether the app or website will tell you if something changes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...who else can see your data on the app or website.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...what parts of your account the app or website can access.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q₂₄ How often do you review what services have access to your Google account?

☐ Never
 ☐ Daily
 ☐ Monthly
☐ Rarely
 ☐ Weekly
 ☐ Yearly

Q₂₅ The following apps are authorized to access various parts of your Google account. Which of these apps would you prefer to keep on your account? Which would you prefer to remove?

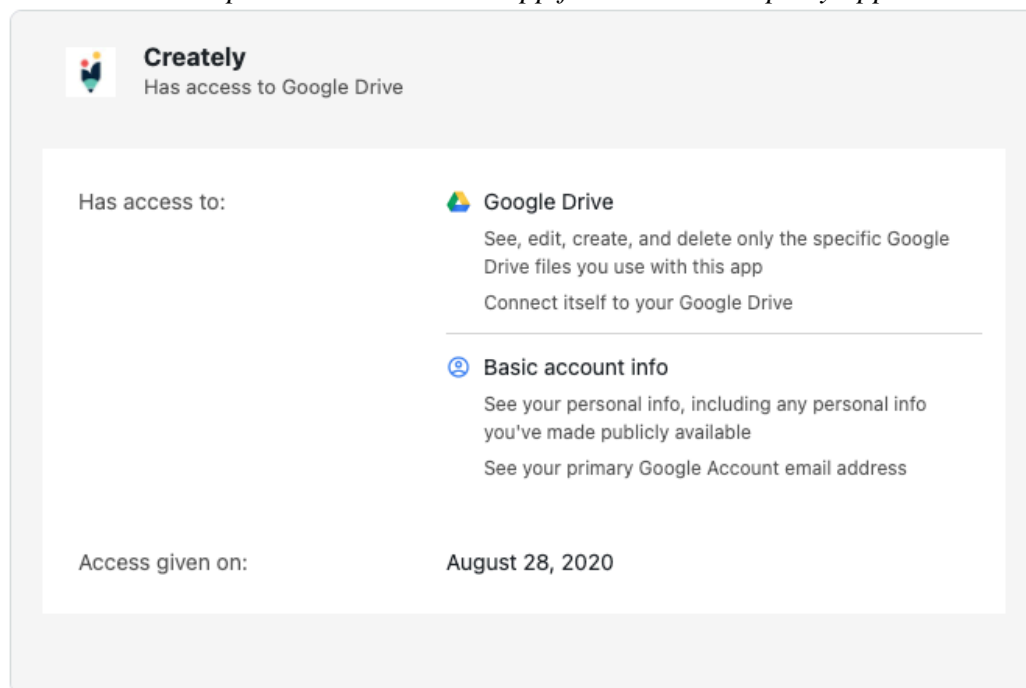
	Keep	Remove	Unsure
[App 1]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[App 2]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[...]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[App n]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q₂₆ Which of these is a sport?

- ☐ Hamburger
 ☐ Basketball
 ☐ Bathroom
 ☐ Skyscraper

[Questions Q₂₇ through Q₂₁₄ are asked repeatedly for the newest app, oldest app, and a random app.]

Participants view a selected app from their third-party apps.



Q₂₇ Do you recall authorizing [App name]?

- ☐ Yes
 ☐ No
 ☐ Unsure

Q₂₈ When was the last time you recall using [App name]?

- ☐ Today
 ☐ In the previous year
☐ In the previous week
 ☐ More than a year ago
☐ In the previous month
 ☐ Unsure

Q₂₉ Prior to seeing the details about [App name], were you aware this app had permission to access parts of your Google account data?

☐ Yes

☐ No

☐ Unsure

Q₂10 Please indicate how strongly you agree or disagree with the following:

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
It's beneficial to me for [App name] to have access to my Google account.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I'm concerned with [App name] having access to my Google account.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I want to change which parts of my Google account that [App name] can access.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q₂11 [App name] holds the following permissions to access parts of your Google account. How ****confident**** are you that you understand what each permission allows the app to do?

	Not confident	Slightly confident	Moderately confident	Confident	Very confident
[App Permission 1]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[App Permission 2]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[...]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[App Permission n]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q₂12 [App name] holds the following permissions to access parts of your Google account. How ****necessary**** do you think each permission is for the app to function in a way that benefits you?

	Not necessary	Slightly necessary	Moderately necessary	Necessary	Very necessary
[App Permission 1]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[App Permission 2]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[...]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[App Permission n]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q₂13 [App name] holds the following permissions to access parts of your Google account. How ****concerned**** are you about the app accessing your account using these permissions?

	Not concerned	Slightly concerned	Moderately concerned	Concerned	Very concerned
[App Permission 1]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[App Permission 2]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[...]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
[App Permission n]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q₂14 Please describe any concerns you have about [App name] holding these permissions. (short answer)

Manage third-party apps & services with access to your Google Account

- By now, you have seen information about apps and websites drawn from the “Apps with access to your account” page on your Google account (similar to the example below).
- You can refer to that page when answering the following questions.

Third-party apps with account access

- You gave these sites and apps access to some of your Google Account data, including info that may be sensitive.

- Remove access for those you no longer trust or use.

Signing in with Google

- You use your Google Account to sign in to these sites and apps.
- They can view your name, email address, and profile picture.

Q₂15 The “Apps with access to your account” page helps me to better understand which third-party apps and websites are linked to my Google account.

- | | |
|--|--------------------------------------|
| <input type="radio"/> Strongly disagree | <input type="radio"/> Agree |
| <input type="radio"/> Disagree | <input type="radio"/> Strongly agree |
| <input type="radio"/> Neither agree nor disagree | |

Q₂16 The “Apps with access to your account” page helps me to better understand what parts of my Google account third-party apps can access.

- | | |
|--|--------------------------------------|
| <input type="radio"/> Strongly disagree | <input type="radio"/> Agree |
| <input type="radio"/> Disagree | <input type="radio"/> Strongly agree |
| <input type="radio"/> Neither agree nor disagree | |

Q₂17 After completing this survey, do you see yourself changing any settings on your “Apps with access to your account” page?

- | | | |
|---------------------------|--------------------------|------------------------------|
| <input type="radio"/> Yes | <input type="radio"/> No | <input type="radio"/> Unsure |
|---------------------------|--------------------------|------------------------------|

Q₂18 In six months do you see yourself reviewing third-party apps from your “Apps with access to your account” page?

- | | | |
|---------------------------|--------------------------|------------------------------|
| <input type="radio"/> Yes | <input type="radio"/> No | <input type="radio"/> Unsure |
|---------------------------|--------------------------|------------------------------|

Q₂19 You have indicated that you would change settings on your “Apps with access to your account” page. Please describe which settings would you change. *[Shown if Q₂17 is “Yes”]* (short answer)

Q₂20 What would you look for when you review the “Apps with access to your account” page in six months? *[Shown if Q₂18 is “Yes”]* (short answer)

Q₂21 What new features (if any) would you like to add to the “Apps with access to your account” page? (short answer)

Q₂22 Suppose Google sent an email reminder to review your “Apps with access to your account” page. How often would you like to be reminded?

- | | |
|------------------------------------|--|
| <input type="radio"/> Once a week | <input type="radio"/> Once every three years |
| <input type="radio"/> Once a month | <input type="radio"/> I do not want to be reminded |
| <input type="radio"/> Once a year | |

Q₂23 Suppose Google required you to reapprove the third-party apps on your Google account. How often would you like to reapprove apps?

- | | |
|------------------------------------|---|
| <input type="radio"/> Once a week | <input type="radio"/> Once every three years |
| <input type="radio"/> Once a month | <input type="radio"/> I do not want to reapprove apps |
| <input type="radio"/> Once a year | |

Q₂24 Which of these is cold?

- ☐ Fire ☐ Summer ☐ Lava ☐ Ice cream

Q₂25 Rather than approve all permissions when installing third-party apps, I would want third-party apps to seek my approval each time they access my Google account.

- ☐ Strongly disagree ☐ Agree
☐ Disagree ☐ Strongly agree
☐ Neither agree nor disagree

Q₂26 I would want to designate specific data (eg. certain emails, individual contacts, particular calendar events) as private and inaccessible to third-party apps.

- ☐ Strongly disagree ☐ Agree
☐ Disagree ☐ Strongly agree
☐ Neither agree nor disagree

B.2 Additional Figures and Tables

Table B.1: Full demographics data of the participants of the first survey.

	Male		Female		Non-binary		PND		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Age	216	50	205	47	11	3	2	0	434	100
18 - 24	58	13	71	16	4	1	0	0	133	31
25 - 34	84	19	67	15	6	1	0	0	157	36
35 - 44	44	10	30	7	1	0	0	0	75	17
45 - 54	19	4	25	6	0	0	0	0	44	10
55 - 64	7	2	8	2	0	0	0	0	15	3
65 or older	4	1	4	1	0	0	0	0	8	2
Prefer not to disclose	0	0	0	0	0	0	2	0	2	0
Highest level of school	216	50	205	47	11	3	2	0	434	100
No schooling completed	0	0	0	0	0	0	0	0	0	0
Some high school	3	1	1	0	0	0	0	0	4	1
High school	20	5	17	4	2	0	0	0	39	9
Some college	48	11	37	9	4	1	0	0	89	21
Trade	4	1	5	1	0	0	0	0	9	2
Associate's Degree	19	4	7	2	1	0	0	0	27	6
Bachelor's Degree	89	21	91	21	3	1	0	0	183	42
Master's Degree	24	6	36	8	1	0	0	0	61	14
Professional degree	4	1	4	1	0	0	0	0	8	2
Doctorate	5	1	6	1	0	0	0	0	11	3
Prefer not to disclose	0	0	1	0	0	0	2	0	3	1
Background	216	50	205	47	11	3	2	0	434	100
Technical	74	17	25	6	1	0	0	0	100	23
Non-Technical	134	31	178	41	10	2	0	0	322	74
Prefer not to disclose	8	2	2	0	0	0	2	0	12	3

Table B.2: Full demographics data of the participants of the second survey.

	Male		Female		Non-binary		PND		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Age	68	46	75	51	4	3	1	1	148	100
18 - 24	18	12	34	23	3	2	0	0	55	37
25 - 34	24	16	21	14	1	1	0	0	46	31
35 - 44	15	10	9	6	0	0	0	0	24	16
45 - 54	7	5	7	5	0	0	0	0	14	9
55 - 64	2	1	3	2	0	0	0	0	5	3
65 or older	2	1	1	1	0	0	0	0	3	2
Prefer not to disclose	0	0	0	0	0	0	1	1	1	1
Highest level of school	68	46	75	51	4	3	1	1	148	100
No schooling completed	0	0	0	0	0	0	0	0	0	0
Some high school	0	0	0	0	0	0	0	0	0	0
High school	7	5	3	2	1	1	0	0	11	7
Some college	20	14	13	9	2	1	0	0	35	24
Trade	1	1	2	1	0	0	0	0	3	2
Associate's Degree	2	1	1	1	0	0	0	0	3	2
Bachelor's Degree	28	19	40	27	1	1	0	0	69	47
Master's Degree	8	5	13	9	0	0	0	0	21	14
Professional degree	1	1	1	1	0	0	0	0	2	1
Doctorate	1	1	1	1	0	0	0	0	2	1
Prefer not to disclose	0	0	1	1	0	0	1	1	2	1
Background	68	46	75	51	4	3	1	1	148	100
Technical	21	14	11	7	0	0	0	0	32	22
Non-Technical	44	30	64	43	4	3	0	0	112	76
Prefer not to disclose	3	2	0	0	0	0	1	1	4	3

Table B.3: The top 26 (4 apps tied for 23rd ranked by authorized count) apps with access, the categories for the requested permissions, the average number of days authorized to the participants' Google account, and the number of permissions each app requested.

App	Permission Categories	Number Authorized	Avg Num of Days Authorized	Number of Permissions
Dropbox	Basic account info, Google Contacts	40	502	4
SAMSUNG Account	Additional access, Basic account info	38	63	5
Shop	Basic account info, Gmail	32	37	3
Nest	Additional access, Basic account info	31	265	3
Windows	Basic account info, Gmail, Google Calendar, Google Contacts	29	397	8
Microsoft SwiftKey	Additional access, Basic account info	27	102	3
macOS	Basic account info, Gmail, Google Calendar, Google Contacts, Google Hangouts	19	452	6
Zoom	Basic account info, Google Calendar, Google Contacts	16	158	5
WhatsApp Messenger	Google Drive	13	663	2
Rakuten Cash Back	Basic account info, Gmail	13	268	3
Pokémon GO	Additional access, Basic account info	10	307	4
Microsoft apps & services	Additional access, Basic account info, Gmail, Google Calendar, Google Contacts, Google Drive	10	290	8
Samsung Email	Basic account info, Gmail	9	1,014	4
Paribus	Basic account info, Gmail	9	390	8
Unroll.me	Basic account info, Gmail, Google Contacts	8	1,389	5
Clash Royale	Google Play	8	644	1
Google Nest Hub	Basic account info	8	287	1
PlayStation Network	Basic account info, YouTube	7	573	5
Shop	Basic account info, Gmail	7	441	5
Chromecast	Basic account info	7	225	1
Lumin PDF	Basic account info, Google Drive	7	164	5
DocHub	Basic account info, Gmail, Google Contacts, Google Drive	7	86	6
IFTTT	Basic account info	6	1,130	1
Sweatcoin	Basic account info	6	486	2
YouTube on Xbox Live	YouTube	6	420	2
Fetch Rewards	Basic account info, Gmail	6	159	3

Table B.4: The top 25 authorized SSO ranked by the number of authorized accesses to participants' Google accounts, and the number of permissions each SSO requested.

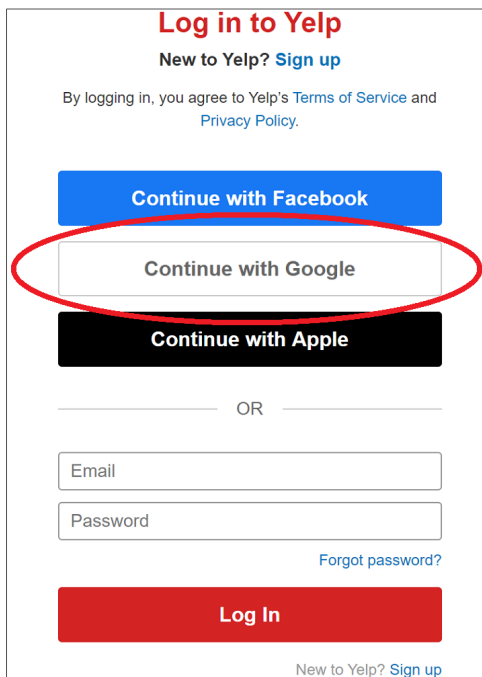
SSO	Permission Categories	Number Authorized	Number of Permissions
Movies Anywhere	Basic account info, Google Play	167	3
Honey	Basic account info	80	2
Amazon Alexa	Basic account info, Gmail, Google Calendar	72	4
Quora	Basic account info, Google Contacts	71	4
Adobe	Basic account info	63	2
Reddit	Basic account info	61	3
Microsoft apps & services	Gmail, Google Calendar, Google Contacts, Google Drive	58	8
Pinterest	Basic account info	58	2
Windows	Basic account info, Gmail, Google Calendar, Google Contacts	53	8
Glassdoor	Additional access, Basic account info	52	3
The New York Times	Basic account info	49	2
Doordash	Basic account info	46	2
Spotify	Basic account info	44	2
macOS	Basic account info, Gmail, Google Calendar, Google Contacts, Google Hangouts	43	6
Quizlet	Basic account info	43	2
Dropbox	Basic account info, Google Contacts	41	4
SAMSUNG Account	Additional access, Basic account info	38	5
Zoom	Basic account info, Google Calendar, Google Contacts	35	5
Shop	Basic account info, Gmail	32	3
Nest	Additional access, Basic account info	31	3
paypal.com	Basic account info	31	2
Adobe Acrobat Reader	Basic account info, Google Drive	30	5
Best Buy App	Basic account info	30	2
Shop	Basic account info, Gmail	30	3
SoundCloud	Basic account info	29	2

Table B.5: The top 28 (5-way tie for permissions count of 8) most requested permissions by third-party apps with authorized access to participants' Google accounts, the permission category, and a count of how many of each permissions were authorized.

Count	Category	Permission
223	Google Play	Create, edit, and delete your Google Play Games activity
189	Basic account info	See your primary Google Account email address
177	Basic account info	See your personal info, including any personal info you've made publicly available
71	Google Drive	See, create, and delete its own configuration data in your Google Drive
44	Basic account info	Associate you with your personal info on Google
28	Google Drive	See, edit, create, and delete only the specific Google Drive files you use with this app
27	Google Drive	See, edit, create, and delete all of your Google Drive files
27	Gmail	Read, compose, send, and permanently delete all your email from Gmail
26	Google Contacts	See, edit, download, and permanently delete your contacts
24	Google Calendar	See, edit, share, and permanently delete all the calendars you can access using Google Calendar
18	Google Contacts	See and download your contacts
16	Basic account info	Full account access
14	Additional access	See and download your exact date of birth
13	Additional access	Display and run third-party web content in prompts and sidebars inside Google applications
12	Gmail	View your email messages and settings
12	Google Drive	See and download all your Google Drive files
11	Additional access	See your age group
11	Additional access	Use Google Fit to see and store your physical activity data
10	Google Docs	See, edit, create, and delete your spreadsheets in Google Drive
10	Google Drive	Connect itself to your Google Drive
10	Additional access	Connect to an external service
9	Additional access	See your language preferences
9	Additional access	See and add to your Google Fit physical activity data
8	Additional access	See and add info about your body measurements and heart rate to Google Fit
8	Gmail	Send email on your behalf
8	Google Docs	See, create, and edit all Google Docs documents you have access to
8	Additional access	See and download all of your Google Account email addresses
8	YouTube	Manage your YouTube account

Table B.6: The top 25 most requested permissions by SSO with authorized access to participants' Google accounts ranked by the number of each permission authorized, and the permission category.

Count	Category	Permission
976	Basic account info	See your primary Google Account email address
938	Basic account info	See your personal info, including any personal info you've made publicly available
88	Basic account info	Associate you with your personal info on Google
45	Google Play	Create, edit, and delete your Google Play Games activity
27	Google Drive	See, edit, create, and delete only the specific Google Drive files you use with this app
25	Gmail	Read, compose, send, and permanently delete all your email from Gmail
25	Google Calendar	See, edit, share, and permanently delete all the calendars you can access using Google Calendar
24	Google Contacts	See, edit, download, and permanently delete your contacts
20	Google Contacts	See and download your contacts
20	Google Drive	See, edit, create, and delete all of your Google Drive files
16	Google Drive	See, create, and delete its own configuration data in your Google Drive
14	Additional access	See and download your exact date of birth
12	Gmail	View your email messages and settings
11	Additional access	See your age group
11	Google Drive	See and download all your Google Drive files
10	Google Docs	See, edit, create, and delete your spreadsheets in Google Drive
10	Google Drive	Connect itself to your Google Drive
10	Additional access	Display and run third-party web content in prompts and sidebars inside Google applications
9	Additional access	See your language preferences
9	Gmail	Send email on your behalf
9	Additional access	See and add info about your body measurements and heart rate to Google Fit
9	Additional access	See and add to your Google Fit physical activity data
9	Additional access	Connect to an external service
8	Additional access	See and download all of your Google Account email addresses
8	Additional access	Use Google Fit to see and store your physical activity data



Log in to Yelp
New to Yelp? [Sign up](#)

By logging in, you agree to Yelp's [Terms of Service](#) and [Privacy Policy](#).

[Continue with Facebook](#)

[Continue with Google](#)

[Continue with Apple](#)

OR

Email

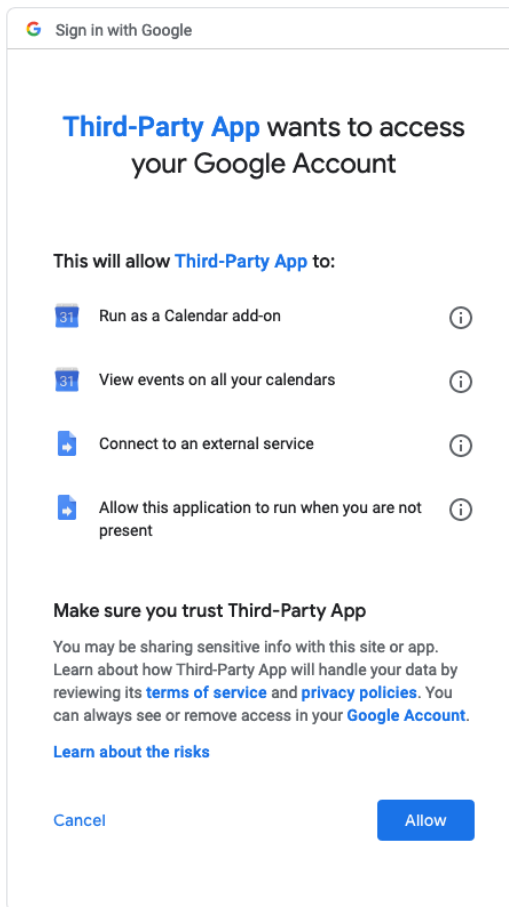
Password

[Forgot password?](#)

[Log In](#)

New to Yelp? [Sign up](#)









(a) Google account SSO prompt.



Sign in with Google

Third-Party App wants to access your Google Account

This will allow **Third-Party App** to:

-  Run as a Calendar add-on 
-  View events on all your calendars 
-  Connect to an external service 
-  Allow this application to run when you are not present 

Make sure you trust Third-Party App

You may be sharing sensitive info with this site or app. Learn about how Third-Party App will handle your data by reviewing its [terms of service](#) and [privacy policies](#). You can always see or remove access in your [Google Account](#).

[Learn about the risks](#)

[Cancel](#) [Allow](#)

(b) Third-party app access consent dialog.

Figure B.1: Google account access authorizations.

← Apps with access to your account

Third-party apps with account access

You gave these sites and apps access to some of your Google Account data, including info that may be sensitive. Remove access for those you no longer trust or use. [Learn about the risks](#)

**Creately**

Has access to Google Drive

**Mozilla Thunderbird Email**

Has access to Gmail

**Zoom**

Has access to Google Calendar

Signing in with Google

You use your Google Account to sign in to these sites and apps. They can view your name, email address, and profile picture. [Learn more](#)


Google Account sign-in prompts

Allow Google to offer a faster way to sign in with your Google Account on supported third-party sites

**Creately****The New York Times**


Google may also have access to some of your third-party accounts. [Learn more](#) about how to manage those connections.

Figure B.2: Google’s “Apps with access to your account” page.

**Dropbox for Gmail**
Has access to Gmail, Google Contacts

REMOVE ACCESS


Has access to:

 Gmail


View your email messages when the add-on is running

Run as a Gmail add-on

Manage drafts and send emails when you interact with the add-on


 Google Contacts

See and download your contacts

 Basic account info


See your personal info, including any personal info you've made publicly available

See your primary Google Account email address

 Additional access

View your country, language, and timezone

Connect to an external service

Homepage: 

https://www.dropbox.com/gmail_integration

Access given on:

August 28, 2020

See something suspicious? [Report this app](#)

Figure B.3: Dropbox for Gmail third-party app as displayed on Google’s “Apps with access to your account” page.

B.3 Qualitative Codebook

- **unconcerned (203)**

permissions-necessary (18), low-permission-level (6), use-often (2), trust-google (1), assume-secure (1), other-primary-email (1)

- **concerned (169)**

permissions (77), personal-data (59), unnecessary-access-to-data (39), misuse (18), data-leak (7), permissions-after-deleted-app (6), privacy (4), tracking (3), permissions-more-permissions-than-described (1), failure-to-update (1), access-when-not-using-app (1)

- **review-app-access (140)**

unused-apps (44), new-apps-with-access (22), what-data-is-accessible (16), do-not-remember-authorizing (12), permissions (10), how-much-access-allowed (8), accidentally-added (7), necessary-permissions-only (7), unfamiliar-apps (6), account-login (4), privacy (4), permissions-changed (4), how-long-access (3), suspicious-apps (3), unauthorized-apps (2), most-used (2), all (1), unauthorized-permissions (1), unnecessary-access (1), specific-app (1), full-account-access (1)

- **remove-app-access (99)**

unused-apps (57), some-apps (11), specific-app (9), unfamiliar-apps (7), all (6), apps-with-account-access (4), apps-with-google-drive-access (2)

- **security (96)**

information (9), misuse (7), information-leak (6), information-theft (3), website (1), app-owner (1), sign-in (1)

- **ease-of-use (94)**

account-creation (13), access-removal (5), sign-in-process (4), fills-in-information (2), how-long (1), already-connected (1), simple (1)

- **unknowns (92)**

information-access (35), use-of-information (15), how-data-used (10), no-unwanted-email (7), what-permission-allows (6), do-they-keep-information (3), email-access (3), why-access-needed (3), why-permissions-necessary (2), why-information-needed (2), who-can-access-data (2), associated-with-personal-info (2), payments (1), can-i-delete-my-data (1), google-guideline-enforcement (1), will-it-notify-on-data-delete (1), how-long-data-stored (1)

- **personal-information (89)**

what-type-accessible-collected (21), how-used (15), amount-collected (9), sensitive (8), is-shared (5), sharing (4), account (3), email (3), calendar (1), kept-confidential (1), photos (1), schedule (1), account-password (1), accounts (1)

- **none (86)**
- **permissions (85)**
necessary (26), access (20), what-information-accessible (13), email-access (5), correct (3), contacts (3), calendar (2), modify-data (2), media (2), location (2), contact-info (1), local-files (1), google-drive (1), webcam (1), alternate-app-with-fewer (1), browsing-history (1)
- **privacy (76)**
data-protection (2), data-used-for-advertising (1), privacy-policy (1)
- **more-transparency (61)**
app-usage-log (13), detailed-permission-explanation (12), when-access-authorized (8), misuse-reports (6), what-data-accessible (5), data-access-logs (5), what-app-needs-to-work (3), what-parts-of-account-accessible (3), which-app-accessed-the-most (1), information-about-security (1), how-data-used (1), company-providing-app (1), how-access-was-authorized (1), terms-and-conditions (1), location-of-app-access (1), listing-of-sites-indirectly-accessing (1), vulnerability-alert (1), permission-usage-log (1), information-about-privacy (1), how-to-remove-information (1)
- **is-trusted (56)**
app (26), website (13), company (5), developer (2)
- **utility (45)**
synch-info (13), file-transfer (6), scan-documents (3), email (3), screenshot (2), edit-pdf (2), install (1), webcam (1), zip-extractor (1), manage-apps (1)
- **other-options (43)**
facebook (14), apple (2), use-without-sign-in (1)
- **nothing (41)**
- **calendar (38)**
- **specific-app-or-service (37)**
zoom (7), google-drive (4), google-docs (2), youtube (1), only-office (1), calendly (1), grammarly (1), google-audio-player (1), snapchat (1), gleam (1), twitter (1), honey (1), slack (1), movies-anywhere (1), plex (1), paypal (1), Doodle (1), google-navigation (1), peloton (1), yelp (1), headspace (1), mcdonalds (1), nest (1), amazon (1), google-play (1), calendars-5 (1), todoist (1), linked-in (1), google-sheets (1)
- **trust (36)**
website (13), app (11), company (5), google (3), site (1), no-unwanted-email (1)
- **is-app-useful (31)**

- **change-permissions (30)**
limit-access (8), remove-unnecessary (4), contacts (4), account-info (2), specific-app (2), email (2), unused-apps (2), calendar (1), delete-files (1), storage (1), sharing (1), google-drive (1), change-files (1)
- **tradeoff (30)**
useful (11), sharing-information (4), frequent-use (1), security (1), account-creation (1), length-of-use (1), frequent-use (1), another-account (1)
- **sign-in (27)**
- **infrequent-use (27)**
- **do-not-recall-authorizing (24)**
permission (6), app (4)
- **app-beneficial (22)**
- **review-changes (22)**
changed-without-notification (2)
- **trust-app (22)**
- **safety-of-app (21)**
- **remember-login (20)**
- **notifications (19)**
reminders (5), upon-changes (2), to-review (2), remove-access-after-app-removed (1), data-breach (1), when-account-accessed (1), email (1), unused-apps (1)
- **gaming (19)**
- **contacts (18)**
- **easier-access-removal (17)**
- **improved-user-interface (17)**
easier-to-access-page (4), sort-apps-by-permission-type (3), personalization (1), order-by-date-authorized (1), edit-preferences-directly (1), color-code-based-on-level-of-access (1), add-app-description-text (1), specific-icons (1), deletion-suggestions (1), filter-by-authorization-date (1), opt-out-option (1)
- **no-recall (15)**
- **personal-info (15)**
health-data (2), location (1), age (1)

- **necessary (15)**
- **shopping (14)**
- **wants-to-protect-personal-data (14)**
leaks (1)
- **unsure (14)**
- **convenience (12)**
account-creation (4)
- **transfer-of-trust (11)**
from-google (10)
- **distrust (11)**
no-unwanted-email (5), google (5), app (1), company (1)
- **fewer-accounts (10)**
- **permission-level-control (9)**
remove-individual-permissions (4), set-some-permissions-never-use (1)
- **unwanted-emails (7)**
- **ability-to-limit-access (7)**
temporarily-block-app-access (2), minimum-necessary (1), remove-access-if-not-used (1), temporarily-allow-app-access (1), time-limits (1)
- **will-remove-app-access (6)**
- **review-how-trusted-app-is (5)**
- **photos (4)**
- **comfort-with-app-or-service (4)**
access-to-information (4)
- **work-related (4)**
- **trade-off-for-convenience (4)**
- **ability-to-secure-data (3)**
limit-what-data-can-be-accessed (1)
- **require-reauthorization (3)**
yearly (1), period-of-time (1), unused-apps (1)

- **will-review-app-access (3)**
- **uncertain (2)**
- **use-separate-account-for-app-access (2)**
- **assess-risks (2)**
- **resigned (2)**
- **app-not-beneficial (2)**
- **education (1)**
- **change-sso-access (1)**
- **remove-app-accessreview-app-access (1)**
- **using-an-alternate-account (1)**
- **social-media (1)**
- **online-reviews (1)**
- **does-not-know-how-to-remove (1)**
- **review-what-data-used (1)**
- **user-reviews (1)**
- **apps-with-access-page-useful (1)**
- **location-tracking (1)**
- **cost (1)**
- **access-to-mobile-device (1)**

Appendix C: Longitudinal Analysis of Privacy Labels

C.1 Additional Figures and Tables

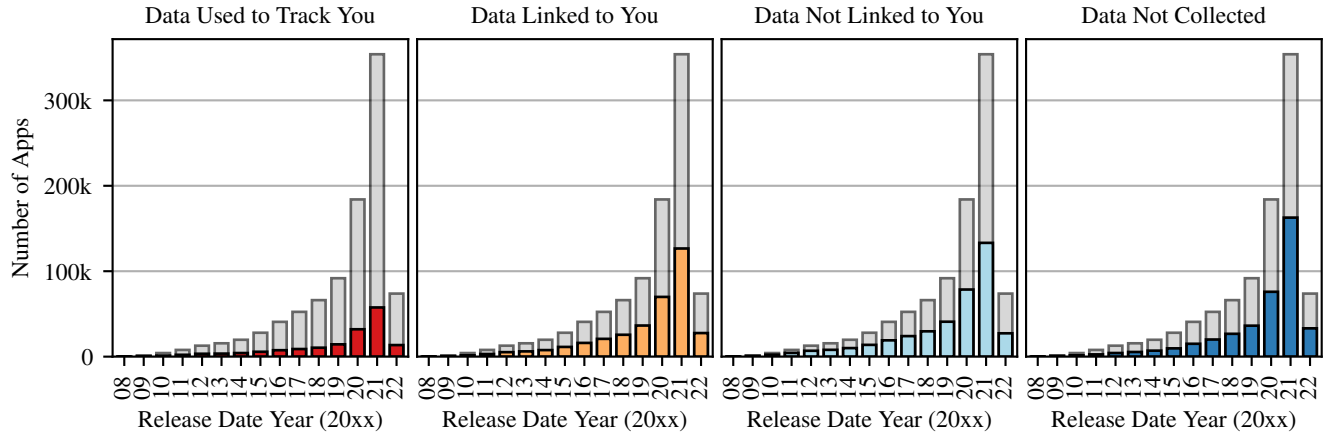


Figure C.1: The number of apps released during a given year for each of the four *Privacy Types*. The gray bars show the total number of apps with privacy labels released in that year. The collection window includes apps through March 17, 2022.

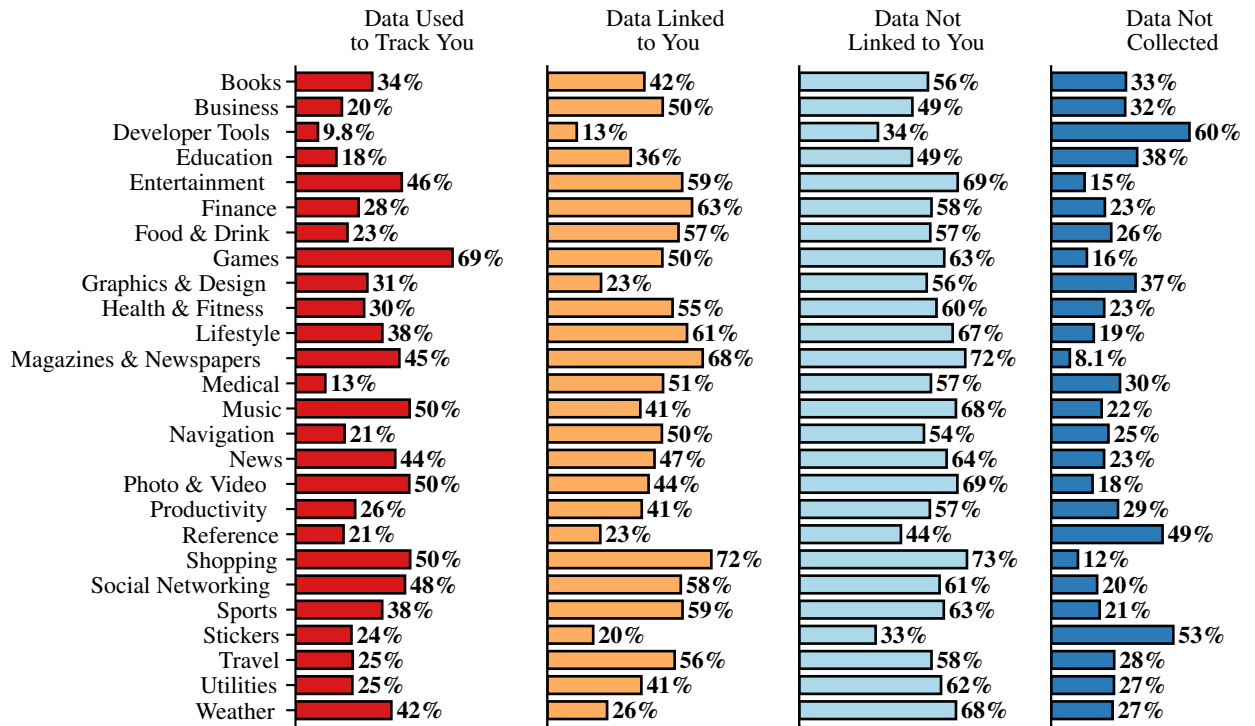


Figure C.2: The ratios of top apps in app store genres for each of the four *Privacy Types*. The denominator is the number of apps with the designated app store genre that have a privacy label. This includes only apps placed in the top in genre categories.

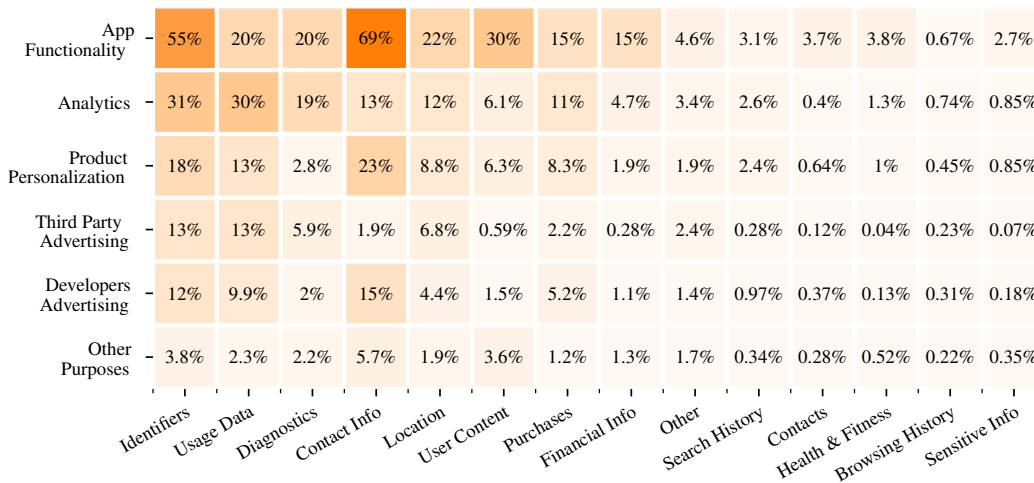


Figure C.3: The ratios of of *Data Categories* by the reported *Purpose* for the *Data Linked to You Privacy Type*.



Figure C.4: The ratios of of *Data Categories* by the reported *Purpose* for the *Data Not Linked to You Privacy Type*.

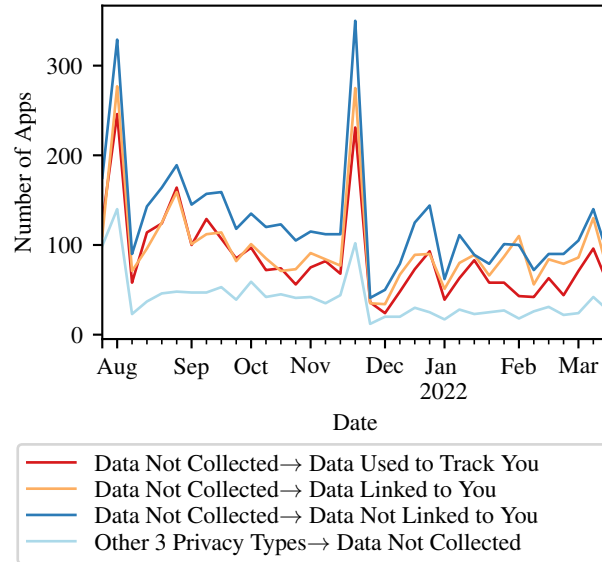


Figure C.5: *Posture* shifts to-and-from *Data Not Collected*. Spikes observed in August and November appear to be due as a result of store-wide delays in updates [12].

Table C.1: Privacy Type Shifts in correlation with version updates.

Initial Posture	New Posture	Without Version Update	With Version Update
Data Not Collected	Data Not Linked to You	1,769	2,629
	Data Linked to You	1,433	1,960
	Data Used to Track You	1,066	1,918
Data Not Linked to You	Data Not Collected	281	685
	Data Linked to You	416	769
	Data Used to Track You	219	364
Data Linked to You	Data Not Collected	246	572
	Data Not Linked to You	586	681
	Data Used to Track You	286	246
Data Used to Track You	Data Not Collected	286	815
	Data Not Linked to You	1,322	2,585
	Data Linked to You	1,353	2,701

Table C.2: Observed shifts in privacy types associated with each data category.

Data Category	From Privacy Type	To Privacy Type	Shift Count	Total Shifts	App Genre
Usage Data	Data Not Linked to You	Data Used to Track You	2,489	6,413	Games
	Data Linked to You	Data Used to Track You	946		
	Data Not Linked to You	Data Linked to You	1,205		
	Data Used to Track You	Data Linked to You	183		
	Data Linked to You	Data Not Linked to You	1,042		
	Data Used to Track You	Data Not Linked to You	548		
Identifiers	Data Not Linked to You	Data Used to Track You	1480	5,904	Games
	Data Linked to You	Data Used to Track You	1,742		
	Data Not Linked to You	Data Linked to You	916		
	Data Used to Track You	Data Linked to You	232		
	Data Linked to You	Data Not Linked to You	1,016		
	Data Used to Track You	Data Not Linked to You	518		
Diagnostics	Data Not Linked to You	Data Used to Track You	590	2,177	Games
	Data Linked to You	Data Used to Track You	154		
	Data Not Linked to You	Data Linked to You	666		
	Data Used to Track You	Data Linked to You	55		
	Data Linked to You	Data Not Linked to You	468		
	Data Used to Track You	Data Not Linked to You	244		
Location	Data Not Linked to You	Data Used to Track You	405	2,013	Games
	Data Linked to You	Data Used to Track You	241		
	Data Not Linked to You	Data Linked to You	398		
	Data Used to Track You	Data Linked to You	47		
	Data Linked to You	Data Not Linked to You	578		

Table C.2 continued from previous page

Data Category	From Privacy Type	To Privacy Type	Shift Count	Total Shifts	App Genre
Contact Info	Data Used to Track You	Data Not Linked to You	344	1,972	Shopping
	Data Not Linked to You	Data Used to Track You	170		
	Data Linked to You	Data Used to Track You	581		
	Data Not Linked to You	Data Linked to You	415		
	Data Used to Track You	Data Linked to You	44		
	Data Linked to You	Data Not Linked to You	507		
	Data Used to Track You	Data Not Linked to You	255		
	Data Not Linked to You	Data Used to Track You	62		
User Content	Data Linked to You	Data Used to Track You	112	825	Business
	Data Not Linked to You	Data Linked to You	238		
	Data Used to Track You	Data Linked to You	6		
	Data Linked to You	Data Not Linked to You	311		
	Data Used to Track You	Data Not Linked to You	96		
	Data Not Linked to You	Data Used to Track You	62		
	Data Linked to You	Data Used to Track You	148		
	Data Not Linked to You	Data Linked to You	82		
Purchases	Data Used to Track You	Data Linked to You	13	457	Games
	Data Linked to You	Data Not Linked to You	103		
	Data Used to Track You	Data Not Linked to You	49		
	Data Not Linked to You	Data Used to Track You	41		
	Data Linked to You	Data Used to Track You	17		
	Data Not Linked to You	Data Linked to You	190		
	Data Used to Track You	Data Linked to You	4		
	Data Linked to You	Data Not Linked to You	107		
Search History				391	Shopping

Table C.2 continued from previous page

Data Category	From Privacy Type	To Privacy Type	Shift Count	Total Shifts	App Genre
Other	Data Used to Track You	Data Not Linked to You	32	253	Games Business Health & Fitness
	Data Not Linked to You	Data Used to Track You	53		
	Data Linked to You	Data Used to Track You	32		
	Data Not Linked to You	Data Linked to You	37		
	Data Used to Track You	Data Linked to You	4		
	Data Linked to You	Data Not Linked to You	76		
	Data Used to Track You	Data Not Linked to You	51		
Financial Info	Data Not Linked to You	Data Used to Track You	6	154	Finance
	Data Linked to You	Data Used to Track You	18		
	Data Not Linked to You	Data Linked to You	38		
	Data Used to Track You	Data Linked to You	2		
	Data Linked to You	Data Not Linked to You	65		
	Data Used to Track You	Data Not Linked to You	25		
Browsing History	Data Not Linked to You	Data Used to Track You	37	152	Shopping
	Data Linked to You	Data Used to Track You	12		
	Data Not Linked to You	Data Linked to You	12		
	Data Used to Track You	Data Linked to You	7		
	Data Linked to You	Data Not Linked to You	63		
	Data Used to Track You	Data Not Linked to You	21		
Contacts	Data Not Linked to You	Data Used to Track You	5	82	Finance
	Data Linked to You	Data Used to Track You	2		
	Data Not Linked to You	Data Linked to You	21		
	Data Used to Track You	Data Linked to You	1		
	Data Linked to You	Data Not Linked to You	38		

Table C.2 continued from previous page

Data Category	From Privacy Type	To Privacy Type	Shift Count	Total Shifts	App Genre
Health & Fitness	Data Used to Track You	Data Not Linked to You	15	52	Medical and Health & Fitness
	Data Not Linked to You	Data Used to Track You	3		
	Data Linked to You	Data Used to Track You	2		
	Data Not Linked to You	Data Linked to You	25		
	Data Used to Track You	Data Linked to You	2		
	Data Used to Track You	Data Not Linked to You	2		
	Data Linked to You	Data Not Linked to You	18		
Sensitive Info	Data Not Linked to You	Data Used to Track You	4	39	Health & Fitness
	Data Linked to You	Data Used to Track You	4		
	Data Not Linked to You	Data Linked to You	13		
	Data Used to Track You	Data Linked to You	2		
	Data Linked to You	Data Not Linked to You	13		
	Data Used to Track You	Data Not Linked to You	3		

Table C.3: Observed shifts in purposes associated with each data category.

Data Category	From Purpose	To Purpose	Count	Total Shifts
Identifiers	App Functionality	Analytics	777	6, 600
	Analytics	App Functionality	318	
	App Functionality	Third Party Advertising	642	
	Third Party Advertising	App Functionality	147	
	App Functionality	Product Personalization	569	
	Product Personalization	App Functionality	71	
	App Functionality	Developers Advertising	559	
	Developers Advertising	App Functionality	87	
	App Functionality	Other Purposes	176	
	Other Purposes	App Functionality	58	
	Analytics	Third Party Advertising	494	
	Third Party Advertising	Analytics	329	
	Analytics	Product Personalization	450	
	Product Personalization	Analytics	147	
	Analytics	Developers Advertising	431	
	Developers Advertising	Analytics	73	
	Analytics	Other Purposes	109	
	Other Purposes	Analytics	47	
	Third Party Advertising	Product Personalization	118	
	Product Personalization	Third Party Advertising	198	
	Third Party Advertising	Developers Advertising	109	
	Developers Advertising	Third Party Advertising	123	
	Third Party Advertising	Other Purposes	39	
	Other Purposes	Third Party Advertising	33	
	Product Personalization	Developers Advertising	188	
	Developers Advertising	Product Personalization	147	
	Product Personalization	Other Purposes	66	
	Other Purposes	Product Personalization	23	
	Developers Advertising	Other Purposes	33	
	Other Purposes	Developers Advertising	39	
	App Functionality	Analytics	371	
	Analytics	App Functionality	539	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
Usage Data	App Functionality	Third Party Advertising	443	6,206
	Third Party Advertising	App Functionality	143	
	App Functionality	Product Personalization	312	
	Product Personalization	App Functionality	91	
	App Functionality	Developers Advertising	306	
	Developers Advertising	App Functionality	94	
	App Functionality	Other Purposes	110	
	Other Purposes	App Functionality	56	
	Analytics	Third Party Advertising	1,040	
	Third Party Advertising	Analytics	204	
	Analytics	Product Personalization	439	
	Product Personalization	Analytics	62	
	Analytics	Developers Advertising	588	
	Developers Advertising	Analytics	37	
	Analytics	Other Purposes	151	
	Other Purposes	Analytics	111	
	Third Party Advertising	Product Personalization	107	
	Product Personalization	Third Party Advertising	213	
	Third Party Advertising	Developers Advertising	124	
	Developers Advertising	Third Party Advertising	145	
Usage Data	Third Party Advertising	Other Purposes	64	
	Other Purposes	Third Party Advertising	41	
	Product Personalization	Developers Advertising	131	
	Developers Advertising	Product Personalization	114	
	Product Personalization	Other Purposes	61	
	Other Purposes	Product Personalization	25	
	Developers Advertising	Other Purposes	48	
	Other Purposes	Developers Advertising	36	
	App Functionality	Analytics	454	
	Analytics	App Functionality	413	
	App Functionality	Third Party Advertising	155	
	Third Party Advertising	App Functionality	67	
	App Functionality	Product Personalization	144	
	Product Personalization	App Functionality	13	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
Diagnostics	App Functionality	Developers Advertising	49	2,181
	Developers Advertising	App Functionality	15	
	App Functionality	Other Purposes	68	
	Other Purposes	App Functionality	123	
	Analytics	Third Party Advertising	170	
	Third Party Advertising	Analytics	63	
	Analytics	Product Personalization	88	
	Product Personalization	Analytics	22	
	Analytics	Developers Advertising	65	
	Developers Advertising	Analytics	10	
	Analytics	Other Purposes	90	
	Other Purposes	Analytics	58	
	Third Party Advertising	Product Personalization	8	
	Product Personalization	Third Party Advertising	19	
	Third Party Advertising	Developers Advertising	8	
	Developers Advertising	Third Party Advertising	6	
	Third Party Advertising	Other Purposes	11	
	Other Purposes	Third Party Advertising	13	
	Product Personalization	Developers Advertising	11	
	Developers Advertising	Product Personalization	9	
	Product Personalization	Other Purposes	3	
	Other Purposes	Product Personalization	13	
	Developers Advertising	Other Purposes	4	
	Other Purposes	Developers Advertising	9	
Contact Info	App Functionality	Analytics	307	2,409
	Analytics	App Functionality	80	
	App Functionality	Third Party Advertising	118	
	Third Party Advertising	App Functionality	14	
	App Functionality	Product Personalization	339	
	Product Personalization	App Functionality	52	
	App Functionality	Developers Advertising	298	
	Developers Advertising	App Functionality	52	
	App Functionality	Other Purposes	139	
	Other Purposes	App Functionality	85	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
Contact Info	Analytics	Third Party Advertising	59	2,409
	Third Party Advertising	Analytics	8	
	Analytics	Product Personalization	84	
	Product Personalization	Analytics	97	
	Analytics	Developers Advertising	99	
	Developers Advertising	Analytics	53	
	Analytics	Other Purposes	44	
	Other Purposes	Analytics	44	
	Third Party Advertising	Product Personalization	13	
	Product Personalization	Third Party Advertising	56	
	Third Party Advertising	Developers Advertising	16	
	Developers Advertising	Third Party Advertising	48	
	Third Party Advertising	Other Purposes	11	
	Other Purposes	Third Party Advertising	16	
	Product Personalization	Developers Advertising	85	
	Developers Advertising	Product Personalization	45	
	Product Personalization	Other Purposes	46	
	Other Purposes	Product Personalization	53	
	Developers Advertising	Other Purposes	19	
	Other Purposes	Developers Advertising	29	
	App Functionality	Analytics	178	
	Analytics	App Functionality	220	
	App Functionality	Third Party Advertising	102	
	Third Party Advertising	App Functionality	93	
	App Functionality	Product Personalization	142	
	Product Personalization	App Functionality	86	
	App Functionality	Developers Advertising	95	
	Developers Advertising	App Functionality	82	
	App Functionality	Other Purposes	74	
	Other Purposes	App Functionality	98	
	Analytics	Third Party Advertising	76	
	Third Party Advertising	Analytics	54	
	Analytics	Product Personalization	74	
	Product Personalization	Analytics	61	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
Location	Analytics	Developers Advertising	78	1,944
	Developers Advertising	Analytics	19	
	Analytics	Other Purposes	70	
	Other Purposes	Analytics	19	
	Third Party Adveritising	Product Personalization	27	
	Product Personalization	Third Party Adveritising	45	
	Third Party Adveritising	Developers Advertising	37	
	Developers Advertising	Third Party Adveritising	22	
	Third Party Adveritising	Other Purposes	17	
	Other Purposes	Third Party Adveritising	12	
	Product Personalization	Developers Advertising	53	
	Developers Advertising	Product Personalization	20	
	Product Personalization	Other Purposes	40	
	Other Purposes	Product Personalization	19	
Location	Developers Advertising	Other Purposes	23	1,944
	Other Purposes	Developers Advertising	8	
	App Functionality	Analytics	135	
	Analytics	App Functionality	48	
	App Functionality	Third Party Adveritising	37	
	Third Party Adveritising	App Functionality	9	
	App Functionality	Product Personalization	108	
	Product Personalization	App Functionality	48	
	App Functionality	Developers Advertising	33	
	Developers Advertising	App Functionality	11	
	App Functionality	Other Purposes	69	
	Other Purposes	App Functionality	89	
	Analytics	Third Party Adveritising	11	
	Third Party Adveritising	Analytics	0	
	Analytics	Product Personalization	30	
	Product Personalization	Analytics	25	
	Analytics	Developers Advertising	15	
	Developers Advertising	Analytics	4	
	Analytics	Other Purposes	17	
	Other Purposes	Analytics	23	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
	Third Party Advertising	Product Personalization	1	
	Product Personalization	Third Party Advertising	16	
	Third Party Advertising	Developers Advertising	0	
	Developers Advertising	Third Party Advertising	2	
	Third Party Advertising	Other Purposes	2	
	Other Purposes	Third Party Advertising	10	
	Product Personalization	Developers Advertising	13	
	Developers Advertising	Product Personalization	12	
	Product Personalization	Other Purposes	18	
	Other Purposes	Product Personalization	17	
	Developers Advertising	Other Purposes	6	
	Other Purposes	Developers Advertising	16	
Purchases	App Functionality	Analytics	64	604
	Analytics	App Functionality	34	
	App Functionality	Third Party Advertising	52	
	Third Party Advertising	App Functionality	7	
	App Functionality	Product Personalization	49	
	Product Personalization	App Functionality	17	
	App Functionality	Developers Advertising	52	
	Developers Advertising	App Functionality	13	
	App Functionality	Other Purposes	12	
	Other Purposes	App Functionality	9	
	Analytics	Third Party Advertising	60	
	Third Party Advertising	Analytics	8	
	Analytics	Product Personalization	31	
	Product Personalization	Analytics	16	
	Analytics	Developers Advertising	43	
	Developers Advertising	Analytics	8	
	Analytics	Other Purposes	8	
	Other Purposes	Analytics	7	
	Third Party Advertising	Product Personalization	6	
	Product Personalization	Third Party Advertising	37	
	Third Party Advertising	Developers Advertising	4	
	Developers Advertising	Third Party Advertising	20	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
Purchases	Third Party Advertising	Other Purposes	3	604
	Other Purposes	Third Party Advertising	6	
	Product Personalization	Developers Advertising	14	
	Developers Advertising	Product Personalization	4	
	Product Personalization	Other Purposes	9	
	Other Purposes	Product Personalization	6	
	Developers Advertising	Other Purposes	0	
	Other Purposes	Developers Advertising	5	
	App Functionality	Analytics	21	
	Analytics	App Functionality	8	
	App Functionality	Third Party Advertising	4	
	Third Party Advertising	App Functionality	2	
	App Functionality	Product Personalization	16	
	Product Personalization	App Functionality	6	
	App Functionality	Developers Advertising	13	
	Developers Advertising	App Functionality	3	
	App Functionality	Other Purposes	16	
	Other Purposes	App Functionality	17	
	Analytics	Third Party Advertising	1	
	Third Party Advertising	Analytics	2	
	Analytics	Product Personalization	3	
	Product Personalization	Analytics	3	
	Analytics	Developers Advertising	4	
	Developers Advertising	Analytics	1	
	Analytics	Other Purposes	4	
	Other Purposes	Analytics	5	
	Third Party Advertising	Product Personalization	1	
	Product Personalization	Third Party Advertising	0	
	Third Party Advertising	Developers Advertising	0	
	Developers Advertising	Third Party Advertising	0	
	Third Party Advertising	Other Purposes	1	
	Other Purposes	Third Party Advertising	2	
	Product Personalization	Developers Advertising	4	
	Developers Advertising	Product Personalization	1	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
Other	Product Personalization	Other Purposes	5	284
	Other Purposes	Product Personalization	8	
	Developers Advertising	Other Purposes	1	
	Other Purposes	Developers Advertising	3	
	App Functionality	Analytics	16	
	Analytics	App Functionality	25	
	App Functionality	Third Party Advertising	14	
	Third Party Advertising	App Functionality	9	
	App Functionality	Product Personalization	15	
	Product Personalization	App Functionality	4	
	App Functionality	Developers Advertising	20	
	Developers Advertising	App Functionality	0	
	App Functionality	Other Purposes	16	
	Other Purposes	App Functionality	15	
	Analytics	Third Party Advertising	19	
	Third Party Advertising	Analytics	3	
	Analytics	Product Personalization	10	
	Product Personalization	Analytics	3	
	Analytics	Developers Advertising	22	
	Developers Advertising	Analytics	0	
	Analytics	Other Purposes	12	
	Other Purposes	Analytics	7	
	Third Party Advertising	Product Personalization	8	
	Product Personalization	Third Party Advertising	7	
	Third Party Advertising	Developers Advertising	8	
	Developers Advertising	Third Party Advertising	3	
	Third Party Advertising	Other Purposes	4	
	Other Purposes	Third Party Advertising	13	
	Product Personalization	Developers Advertising	8	
	Developers Advertising	Product Personalization	1	
	Product Personalization	Other Purposes	7	
	Other Purposes	Product Personalization	7	
	Developers Advertising	Other Purposes	3	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
Search History	Other Purposes	Developers Advertising	5	291
	App Functionality	Analytics	28	
	Analytics	App Functionality	44	
	App Functionality	Third Party Advertising	14	
	Third Party Advertising	App Functionality	1	
	App Functionality	Product Personalization	23	
	Product Personalization	App Functionality	14	
	App Functionality	Developers Advertising	11	
	Developers Advertising	App Functionality	5	
	App Functionality	Other Purposes	4	
	Other Purposes	App Functionality	3	
	Analytics	Third Party Advertising	18	
	Third Party Advertising	Analytics	0	
	Analytics	Product Personalization	36	
	Product Personalization	Analytics	15	
	Analytics	Developers Advertising	18	
	Developers Advertising	Analytics	0	
	Analytics	Other Purposes	4	
	Other Purposes	Analytics	4	
	Third Party Advertising	Product Personalization	0	
	Product Personalization	Third Party Advertising	13	
	Third Party Advertising	Developers Advertising	1	
	Developers Advertising	Third Party Advertising	11	
	Third Party Advertising	Other Purposes	0	
	Other Purposes	Third Party Advertising	3	
	Product Personalization	Developers Advertising	8	
	Developers Advertising	Product Personalization	6	
	Product Personalization	Other Purposes	2	
	Other Purposes	Product Personalization	1	
Search History	Developers Advertising	Other Purposes	1	291
	Other Purposes	Developers Advertising	3	
	App Functionality	Analytics	5	
	Analytics	App Functionality	5	
	App Functionality	Third Party Advertising	1	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
Contacts	Third Party Advertising	App Functionality	1	141
	App Functionality	Product Personalization	90	
	Product Personalization	App Functionality	7	
	App Functionality	Developers Advertising	5	
	Developers Advertising	App Functionality	0	
	App Functionality	Other Purposes	5	
	Other Purposes	App Functionality	6	
	Analytics	Third Party Advertising	2	
	Third Party Advertising	Analytics	0	
	Analytics	Product Personalization	1	
	Product Personalization	Analytics	2	
	Analytics	Developers Advertising	2	
	Developers Advertising	Analytics	1	
	Analytics	Other Purposes	1	
	Other Purposes	Analytics	0	
	Third Party Advertising	Product Personalization	0	
	Product Personalization	Third Party Advertising	1	
	Third Party Advertising	Developers Advertising	0	
	Developers Advertising	Third Party Advertising	1	
	Third Party Advertising	Other Purposes	0	
	Other Purposes	Third Party Advertising	1	
	Product Personalization	Developers Advertising	1	
	Developers Advertising	Product Personalization	0	
	Product Personalization	Other Purposes	1	
	Other Purposes	Product Personalization	2	
	Developers Advertising	Other Purposes	0	
	Other Purposes	Developers Advertising	0	
	App Functionality	Analytics	14	
	Analytics	App Functionality	9	
	App Functionality	Third Party Advertising	3	
	Third Party Advertising	App Functionality	0	
	App Functionality	Product Personalization	39	
	Product Personalization	App Functionality	9	
	App Functionality	Developers Advertising	4	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
Health & Fitness	Developers Advertising	App Functionality	0	141
	App Functionality	Other Purposes	1	
	Other Purposes	App Functionality	4	
	Analytics	Third Party Advertising	0	
	Third Party Advertising	Analytics	0	
	Analytics	Product Personalization	43	
	Product Personalization	Analytics	4	
	Analytics	Developers Advertising	1	
	Developers Advertising	Analytics	0	
	Analytics	Other Purposes	2	
	Other Purposes	Analytics	2	
	Third Party Advertising	Product Personalization	0	
	Product Personalization	Third Party Advertising	0	
	Third Party Advertising	Developers Advertising	0	
Health & Fitness	Developers Advertising	Third Party Advertising	0	141
	Third Party Advertising	Other Purposes	0	
	Other Purposes	Third Party Advertising	1	
	Product Personalization	Developers Advertising	1	
	Developers Advertising	Product Personalization	0	
	Product Personalization	Other Purposes	1	
	Other Purposes	Product Personalization	2	
	Developers Advertising	Other Purposes	0	
	Other Purposes	Developers Advertising	1	
	App Functionality	Analytics	22	
	Analytics	App Functionality	8	
	App Functionality	Third Party Advertising	4	
	Third Party Advertising	App Functionality	2	
	App Functionality	Product Personalization	3	
	Product Personalization	App Functionality	1	
	App Functionality	Developers Advertising	21	
	Developers Advertising	App Functionality	4	
	App Functionality	Other Purposes	2	
	Other Purposes	App Functionality	0	
	Analytics	Third Party Advertising	4	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
Browsing History	Third Party Advertising	Analytics	1	104
	Analytics	Product Personalization	4	
	Product Personalization	Analytics	1	
	Analytics	Developers Advertising	5	
	Developers Advertising	Analytics	0	
	Analytics	Other Purposes	4	
	Other Purposes	Analytics	1	
	Third Party Advertising	Product Personalization	1	
	Product Personalization	Third Party Advertising	4	
	Third Party Advertising	Developers Advertising	3	
	Developers Advertising	Third Party Advertising	2	
	Third Party Advertising	Other Purposes	0	
	Other Purposes	Third Party Advertising	1	
	Product Personalization	Developers Advertising	1	
	Developers Advertising	Product Personalization	1	
	Product Personalization	Other Purposes	1	
	Other Purposes	Product Personalization	1	
	Developers Advertising	Other Purposes	1	
	Other Purposes	Developers Advertising	1	
Sensitive Info	App Functionality	Analytics	9	56
	Analytics	App Functionality	3	
	App Functionality	Third Party Advertising	4	
	Third Party Advertising	App Functionality	0	
	App Functionality	Product Personalization	5	
	Product Personalization	App Functionality	1	
	App Functionality	Developers Advertising	2	
	Developers Advertising	App Functionality	1	
	App Functionality	Other Purposes	5	
	Other Purposes	App Functionality	2	
	Analytics	Third Party Advertising	3	
	Third Party Advertising	Analytics	0	
	Analytics	Product Personalization	2	
	Product Personalization	Analytics	4	
	Analytics	Developers Advertising	2	

Table C.3 continued from previous page

Data Category	From Purpose	To Purpose	Count	Total Shifts
Sensitive Info	Developers Advertising	Analytics	1	56
	Analytics	Other Purposes	2	
	Other Purposes	Analytics	1	
	Third Party Adveritising	Product Personalization	0	
	Product Personalization	Third Party Adveritising	3	
	Third Party Adveritising	Developers Advertising	0	
	Developers Advertising	Third Party Adveritising	1	
	Third Party Adveritising	Other Purposes	0	
	Other Purposes	Third Party Adveritising	1	
	Product Personalization	Developers Advertising	1	
	Developers Advertising	Product Personalization	0	
	Product Personalization	Other Purposes	2	
	Other Purposes	Product Personalization	1	
	Developers Advertising	Other Purposes	0	
	Other Purposes	Developers Advertising	0	

Appendix D: Advertisement Inferences

D.1 Survey Instrument

Thank you for your interest in our survey. Your answers are important to us. **Please read the following instructions carefully:** (i) Take your time in reading and answering the questions. (ii) Answer the questions as accurately as possible. (iii) It is okay to say that you don't know an answer.

[A horizontal rule, like below, indicates a new page in the questionnaire.]

S1 Do you have a personal Gmail address (an email address ending in “gmail.com”)?

- ☐ Yes ☐ No

S2 How long do you have that Gmail address?

- ☐ Less than a year ☐ More than five years
☐ One year ☐ I do not have a Gmail address
☐ Three years ☐ Unsure
☐ Five years

S3 Which other Google products do you currently use? (Select all that apply.)

- ☐ Gmail ☐ Google Search ☐ Google Pay
☐ Google Maps ☐ Google Play ☐ Android device
☐ YouTube ☐ Google Drive ☐ None of these
☐ Google Chrome ☐ Google News

[Included only products selected in S3. If “None of these” was selected question was hidden]

S4 How frequently do you use these products?

	Always	Often	Sometimes	Rarely	Never	Unsure
Gmail	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Maps	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
YouTube	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Chrome	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Search	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Play	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Drive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google News	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Google Pay	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Android device	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

S5 How important is using Google products to your Internet experience?

- ☐ Not important ☐ Important
☐ Slightly important ☐ Very important
☐ Moderately important
-

Please install the “Survey Assistant” browser extension.

On this page, we will ask you to install our browser extension, which will assist you during the study. The installation takes only a few steps. Please follow the instructions below.

We will assist you to uninstall the browser extension at the end of the survey.

Google Sign-in

This survey also requires that you login to **your primary Google account** for accessing items in your My Activity and Ad Settings page. My Activity and Ad Settings are services provided by Google to make you see more useful ads and offer a more personalized experience, including faster searches and automatic recommendations. Please use an account with an email address that ends in **@gmail.com**.

Privacy Note: We do not track or store your email address as part of this study, and we will not be able to tie your email address to any results or analysis. The researchers will never see your email address. At no time do the researchers have access to your Google account.

Q1 How aware are you of the amount of information that Google is collecting about your activities online?

- | | |
|--|--|
| <input type="radio"/> Not at all aware | <input type="radio"/> Moderately aware |
| <input type="radio"/> Slightly aware | <input type="radio"/> Extremely aware |
| <input type="radio"/> Somewhat aware | |

Q2 How concerned are you with the amount of information Google is collecting about your activities online?

- | | |
|--|--|
| <input type="radio"/> Not at all concerned | <input type="radio"/> Moderately concerned |
| <input type="radio"/> Slightly concerned | <input type="radio"/> Extremely concerned |
| <input type="radio"/> Somewhat concerned | |

Q2_A Please explain why.

Answer: _____

Q3 How often do you benefit from the amount of information that Google collects about your activities online?

- | | |
|---------------------------------|------------------------------|
| <input type="radio"/> Never | <input type="radio"/> Often |
| <input type="radio"/> Rarely | <input type="radio"/> Always |
| <input type="radio"/> Sometimes | |

Q3_A Please explain why.

Answer: _____

What is Google Ad Settings

The following video briefly introduces you to Google's Ad Settings page. For every account, Google provides a settings page called Ad Settings, which allows you to control whether and how personalized ads are shown to you.

Please watch this 1-minute video before proceeding to the next page:

[YouTube video describing Google's Ad Settings.]



Prepare Survey

Please click the “Prepare Survey” button to proceed to the next page.

[Interactive checklist, which visualized the progress of gathering data from the My Activity and Ad Settings page.]

- Check for browser extension
 - Check for ad interests (show number of found interests)
 - Check for Google Search activities (show number of found Google Search activities)
 - Check for YouTube activities (show number of found YouTube activities)
 - Check for Google Maps activities (show number of found search Google Maps)
-

Task Instructions

Next, we will ask you to assign at least 3 interests you think Google will learn from the shown activity, which we randomly selected from your Google MyActivity page. This task is repeated with up to 9 different activities.

Note: We will store your assigned interests and activities. Please make sure to only assign interests to activities you feel comfortable sharing with us. You can skip any activity by clicking “Skip this Activity”.

Before we start, let us explain how you can assign these interests on the next page.

Task Demo

[Interactive tutorial of the Matching Task user interface.]

1. **The Activity:** On the next page, you will be presented with one of your activities picked from your Google My Activity page.
 2. **Skipping an Activity:** If you feel uncomfortable with the shown activity you can skip the activity by pressing this button.
 3. **Reason for Skipping:** If you choose to skip, please select why you do not like to assign an interest and share the activity with us.
 4. **Assign an Interest:** To assign an interest you think Google will learn from the shown activity above, use this text box to enter your idea.
 5. **List of Interests:** While you type, we will display some suggestions that Google typically applies. You can select one by clicking on it.
 6. **Selected Interests:** Your selected interests will be displayed in the text box and can be removed by pressing on the “X” icon.
 7. **Custom Interests:** If no suggestion fits your idea, you can also enter your own interests.
 8. **Add Custom Interest:** To add a custom interest, please press the plus button or the “Enter” key on your keyboard.
 9. **Finished:** To start your assignment task, please click the Next button to proceed to the next page.
-

Matching Tasks

In the next part of the survey, we will ask you questions about **nine activities** from your My Activity page. The activities are chosen randomly. We do not collect information about that activity as part of this survey. That information remains private, **only accessible to you** and Google. We only note which service the activity is associated with, e. g., “Google search” vs. “YouTube view”, and the date on which it occurred. Further details are not collected as part of this survey.

[View of the matching task for a YouTube Activity.]

YouTube Activity


Apr 20, 2022

YouTube

Watched 2022 Mazda CX-30 | Mazda USA

Watched at 6:46pm

10:46 PM



Skip this Activity

Undo Skip

Seeing the activity above, what interests do you think Google would apply to it? *

Please type or select at least 3 interests.

S Sports Cars

A Autos & Vehicles

A Automotive Industry

M Mazda

Type to search...

+

I(1-9) Seeing the activity above, what interests do you think Google would apply to it?

[This question was repeated up to 9 times with different activities.]

Now that you have worked with activities and interests, we would like to repeat some of the questions from the beginning of the survey.

Q9 How concerned are you with the amount of information Google is collecting about your activities online?

- | | |
|--|--|
| <input type="radio"/> Not at all concerned | <input type="radio"/> Moderately concerned |
| <input type="radio"/> Slightly concerned | <input type="radio"/> Extremely concerned |
| <input type="radio"/> Somewhat concerned | |

Q9_A Please explain why you changed or not changed your assessment.

-

Q10 How often do you benefit from the amount of information that Google collects about your activities online?

- | | |
|---------------------------------|------------------------------|
| <input type="radio"/> Never | <input type="radio"/> Often |
| <input type="radio"/> Rarely | <input type="radio"/> Always |
| <input type="radio"/> Sometimes | |

Q10_A Please explain why you changed or not changed your assessment.

-

























Google Ad Settings

Next, we will show you the interests from your Google Ad Settings page. These interests are assigned to your account by Google.

Note: Please take your time to look through the interests. You may proceed to the next page after 30 seconds.

Presentation of all inferences

[View of the participants inferences from their ad settings page.]

 Apparel	 Autos & Vehicles
 Banking	 Books & Literature
 Business Formation	 Business News
 Business Services	 Camera & Photo Equipment
 Cats	 Celebrities & Entertainment News
 Charity & Philanthropy	 Company Size: Large Employer (250-10k E...
 Competitive Video Gaming	 Computer & Video Games
 Computer Hardware	 Computers & Electronics
 Construction & Maintenance	 Cooking & Recipes
 Cookware & Diningware	 Credit Cards
 Cutlery & Cutting Accessories	 Distributed & Cloud Computing
 DJ Resources & Equipment	 Dogs

Q11 How concerned are you about Google learning about your interests based on your activities online?

- ☐ Not at all concerned
 ☐ Moderately concerned
☐ Slightly concerned
 ☐ Extremely concerned
☐ Somewhat concerned

Q11_A Please explain why.

- Please type in your answer.

Q12 How often do you benefit from Google learning your interests based on your activities online?

- ☐ Never
 ☐ Often
☐ Rarely
 ☐ Always
☐ Sometimes

Q12_A Please explain why.

- Please type in your answer.

Q13 How accurate do you think Google is when learning your interests?

- ☐ Not at all accurate
 ☐ Moderately accurate
☐ Slightly accurate
 ☐ Extremely accurate
☐ Somewhat accurate

Q13_A Please explain why.

- Please type in your answer.

D1 What is your gender? (Optional question)

- ☐ Woman
- ☐ Man
- ☐ Non-binary
- ☐ Prefer not to disclose
- ☐ Prefer to self-describe

D2 What is your age? (Optional question)

- ☐ 18 – 24
- ☐ 25 – 34
- ☐ 35 – 44
- ☐ 45 – 54
- ☐ 55 – 64
- ☐ 65 or older
- ☐ Prefer not to disclose

D3 Are you of Hispanic, Latino, or Spanish origin? (Optional question)

- ☐ No, not of Hispanic, Latino, or Spanish origin,
- ☐ Yes, Mexican, Mexican Am., Chicano,
- ☐ Yes, Puerto Rican,
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin
 - Type, for example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc. _____

D4 What is your race? (Optional question) Please type one or more origins.

- ☐ White
- ☐ Black or African American
- ☐ American Indian or Alaskan Native
- ☐ Chinese
- ☐ Filipino
- ☐ Asian Indian
- ☐ Vietnamese
- ☐ Korean
- ☐ Japanese
- ☐ Other Asian
- ☐ Native Hawaiian
- ☐ Samoan
- ☐ Chamorro
- ☐ Other Pacific Islander

D5 Which of these describes your personal income last year? (Optional question)

- ☐ No income
- ☐ \$1 to \$9,999
- ☐ \$10,000 to \$24,999
- ☐ \$25,000 to \$49,999
- ☐ \$50,000 to \$74,999
- ☐ \$75,000 to \$99,999
- ☐ \$100,000 to \$149,999
- ☐ \$150,000 and greater
- ☐ Prefer not to disclose

D6 What is the highest degree or level of school you have completed? (Optional question)

- ☐ No schooling completed
- ☐ Some high school, no diploma
- ☐ High school graduate, diploma, or equivalent
- ☐ Some college credit, no degree
- ☐ Trade / technical / vocational training
- ☐ Associate degree
- ☐ Bachelor's degree
- ☐ Master's degree
- ☐ Professional degree (e. g., J.D., M.D.)
- ☐ Doctorate degree
- ☐ Prefer not to disclose
- ☐ Other (please specify)

D7 Do you have a major in one of the following fields? (Optional question)

- ☐ Science, technology, engineering, and mathematics (S.T.E.M.)
- ☐ Humanities
- ☐ Law
- ☐ Social science
- ☐ Prefer not to disclose

☐ Other (please specify)

D8 What is your marital status? (Optional question)

- ☐ Married
- ☐ Living as married
- ☐ Divorced
- ☐ Widowed

- ☐ Separated
- ☐ Single, never been married
- ☐ Prefer not to disclose
- ☐ Prefer to self-describe

D9 How many children do you have? (Optional question)

- ☐ 0
- ☐ 1
- ☐ 2–4

- ☐ 4+
- ☐ Prefer not to disclose

D10 Which of the following best represents how you think of yourself? (Optional question)

- ☐ Lesbian or gay
- ☐ Straight, that is, not lesbian or gay
- ☐ Bisexual
- ☐ Something else
- ☐ I don't know the answer
- ☐ Prefer not to disclose
- ☐ Prefer to self-describe

Appendix E: User Perceptions of App-Based Privacy Labels

E.1 Survey Instrument

Thank you for your interest in our survey. Your answers are important to us. **Please read the following instructions carefully:** (i) Take your time in reading and answering the questions. (ii) Answer the questions as accurately as possible. **Definitions:** (i) App: In this survey the word “app” refers to an application found on the Apple App Store that can be installed on your Apple device. (ii) Privacy Label: a short summary of an app’s data collection behavior displayed on the application pages of the Apple App Store.

On the next page we will provide an introduction to this survey.

[A horizontal rule, like below, indicates a new page in the questionnaire.]

Survey Introduction

This survey is designed to investigate your awareness of app privacy labels displayed on the application pages of the Apple App Store. You will answer questions regarding potential app installation decisions and how an app privacy label may impact your thoughts about the app. On the following pages you will be presented with an application and asked questions about this application and its privacy labels. For each of the labels we will ask a set of similar questions, so please pay close attention.

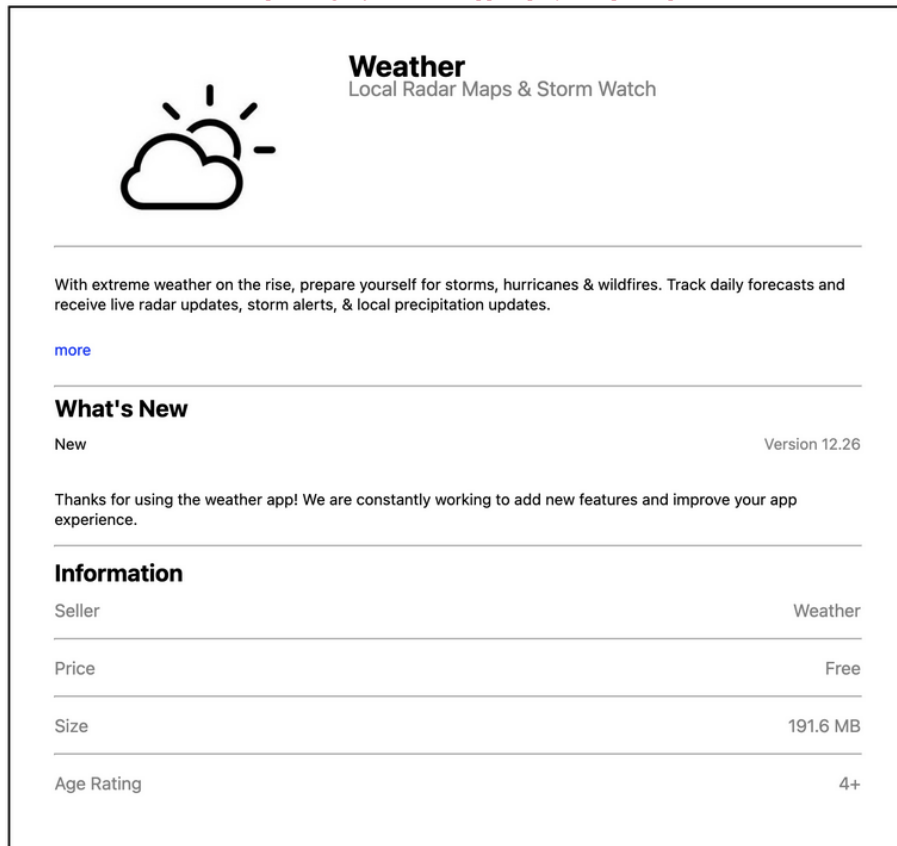
App Related Questions

[Participants are randomly assigned a weather app or a social media app.]

Imagine you are making a decision to install a *[App Name]* app on your phone that was recommended by a friend. The price of the app is within your budget (or it is free) and the features are what you would expect from a *[App Name]* app.

Please review the description for the *[App Name]* app before answering the questions below.

[An example image of a weather app displayed to participants.]



Q1 How concerned are you about the way the *[App Name]* app shown above will collect, store, and use information?

- ☐ Not at all concerned
☐ Slightly concerned
☐ Somewhat concerned

- ☐ Moderately concerned
☐ Very concerned

Q2 What about data collection, storage, and use by the *[App Name]* app makes you feel concerned?

-

Q3 Do you currently have a *[App Name]* app installed on your phone?

- ☐ Yes
 ☐ No

[Included only if Yes selected in Q3.]

Q4 How long have you had this *[App Name]* app? If you have more than one device, answer the question for the one that you have had for the longest time.

- ☐ Less than a month
 ☐ More than a year
☐ Between a month and a year
 ☐ I don't remember

Q5 What were your reasons to install the *[App Name]* app?

-

[Included only if No selected in Q3.]

Q6 Have you ever considered installing a *[App Name]* app on your phone?

- ☐ Yes
 ☐ No

[Included only if Yes selected in Q6.]

Q7 What made you decide not to install the *[App Name]* app?

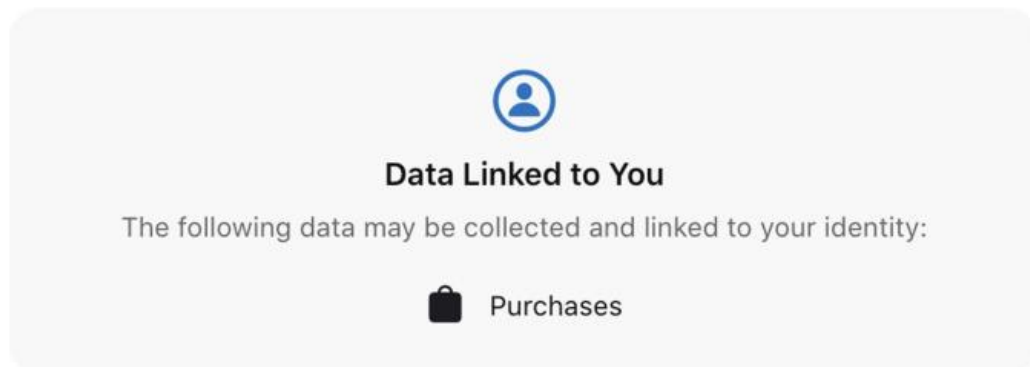
-

Privacy Label Related Questions

[Q8 - Q12 will be asked once per privacy label. The privacy labels are chosen randomly. There will be up to 4 privacy labels shown to each study participant]

Please imagine the following privacy label (a short summary of the app's data collection behavior) was shown on the App Store page of the app when answering the questions below.

[An example image of a privacy label displayed to participants.]



Q8 How confident are you that you know what the label shown above means?

- ☐ Not at all confident
 ☐ Moderately confident
☐ Slightly confident
 ☐ Very confident
☐ Somewhat confident

Q9 I believe the label shown above

- ☐ Strongly decreases the privacy and security risks associated with this specific *[App Name]* app
☐ Slightly decreases the privacy and security risks associated with this specific *[App Name]* app
☐ Does not have any impact on the privacy and security risks associated with this specific *[App Name]* app
☐ Slightly increases the privacy and security risks associated with this specific *[App Name]* app
☐ Strongly increases the privacy and security risks associated with this specific *[App Name]* app

Q10 Please explain why you believe the label (decreases/increases/does not have any impact on) the privacy and security risks associated with this specific app

-

Q11 Assuming you want to install the *[App Name]* app on your phone, knowing that this app has the label shown above would

- ☐ Strongly decrease your willingness to install this app.
☐ Slightly decrease your willingness to install this app.
☐ Not have any impact on your willingness to install this app.
☐ Slightly increase your willingness to install this app.
☐ Strongly increase your willingness to install this app.

Q12 Please explain why knowing that this app has the label (decreases/increases/does not have any impact on) your willingness to install *[App Name]*

-

Demographics

D1 What is your gender?

- ☐ Woman
 ☐ Prefer not to disclose
☐ Man
 ☐ Prefer to self-describe
☐ Non-binary

D2 What is your age?

- ☐ 18 – 24
 ☐ 55 – 64
☐ 25 – 34
 ☐ 65 or older
☐ 35 – 44
 ☐ Prefer not to disclose
☐ 45 – 54

D3 Are you a student?

- ☐ Yes
 ☐ No
 ☐ Prefer not to disclose

D4 What is the highest degree or level of school you have completed?

- ☐ No schooling completed
☐ Some high school, no diploma
☐ High school graduate, diploma, or equivalent
☐ Some college credit, no degree
☐ Trade / technical / vocational training
☐ Associate degree
☐ Bachelor's degree
☐ Master's degree
☐ Professional degree (e. g., J.D., M.D.)
☐ Doctorate degree
☐ Prefer not to disclose
☐ Other (please specify)

D5 Which of the following best describes your educational background or job field?

- ☐ I have an education in, or work in, the field of computer science, computer engineering or IT.
☐ I do not have an education in, nor do I work in, the field of computer science, computer engineering or IT.
☐ Prefer not to disclose