Regulatory Technologies for the Study of Data and Platform Power in the App Economy

Konrad Kollnig

Hertford College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2023

To a future, in which we can prosper freely, and in peace with technology.

Acknowledgements

My foremost thanks go to Sir Nigel Shadbolt. He has been a source of inspiration throughout my DPhil and continuously challenged me to keep track of the bigger picture. He encouraged me to enhance my technical skills relentlessly, to follow non-traditional pathways and challenge the status quo, to put societal impact over the number of publications, and to strive for meaningful connections with others outside of the Oxford bubble. In short, he gave me the freedom and space that is needed to produce transformative research. It has been a true pleasure.

I thank Siddhartha Datta, Pierre Dewitte and Dr Anastasia Shuba. Siddhartha is the most talented young researcher that I know. He combines unique technical expertise with a strong passion for the societal impacts of digital technologies. I am deeply grateful to have met and to have learned from each other. Pierre has seemingly limitless dedication and creativity in using the law to make tech more ethical. He is now even taking some of the most egregious data practices all the way up to the European Court of Justice. In his other life, he might well have become a computer scientist. Anastasia is, by her own admission, a software engineer by day and privacy researcher by night. She taught me the art of conducting and writing technical papers and was available whenever I got stuck. She also showed me that collaborative research can be done in truly good faith and for the cause at hand.

Further, I thank all my colleagues and collaborators, particularly Dr Nitin Agrawal, Jumana Baghabrah, Dr Reuben Binns, Dr Paul-Olivier Dehaye, Anirudh Ekambaranathan, Dr Tess Johnson, Yury Kolotaev, Dr Ulrik Lyngs, Glòria Macià Muñoz, Dr Vince Madai, Daniel Omeiza, Thomas Şerban von Davier, Jake Stein, Claudine Tinsman, Dr Max Van Kleek, Tiffany Ge Wang, Dr Helena Webb and Dr Jun Zhao. Without their time and devotion, this thesis would not have been possible.

Besides spending much time at the Oxford Department of Computer Science, I was fortunate to be a member of the Computers and Law Research Group of the Oxford Faculty of Law and to attend the reading groups and conferences organised by members of that group. I am thankful for thoughtful discussions and exchange with Prof Jeremias Adams-Prassl, Dr Halefom Abraha, Jinghe Fan, Jyothsna Gurumurthy, Dr Václav Janeček, Aislinn Kelly-Lyth, Juliana Mota, Luz Orozco Y Villa, Dr Six Silberman, and Peter Wills.

My thanks also go to the participants of the Venice AI colloquium, jointly organised by Sir Nigel Shadbolt, Lord David Neuberger, and Dr Julie Maxton. It was an honour to meet extraordinary members of Goodenough College London, including The Hon Alice Walpole, The Revd Dr Alan McCormack, Tess Buckley, Alexandra Houston, Dr Karan Mehta, Zheng Hong Sebastian See, Luke Thorburn, and Rachel Venn. The conference provided much valuable input for the last stretch of my DPhil and my career plans.

In the last few months of my DPhil, I was given the unique opportunity to support the ongoing work of the Open Data Institute on Privacy-Enhancing Technologies and thank my colleagues there, especially Calum Inverarity and Anastasia Shteyn.

I thank the users of my TrackerControl app for helping make the app what it has become, building a vibrant community, and translating the app into more than 21 languages.

I am thankful for generous financial support received from Hertford College, the UK Engineering and Physical Sciences Research Council (EPSRC), the Oxford Martin School EWADA Programme, DeepMind, and the University of Oxford.

Lastly, I thank my close friends and family – particularly my mum, dad, and siblings, Connor, Yannick, Sophie, Johannes, and Lena – for their support of me and my work, which was not always easy during the Covid-19 pandemic. Pursuing research can already be isolating in normal times, and the pandemic brought much academic exchange to a standstill. *But now, onward! There's much to do.*

Konrad Kollnig
Hertford College, Oxford
Hilary 2023

Abstract

Tracking, the large-scale collection of data about user behaviour, is commonplace in mobile apps. While some see tracking as a necessary evil to making apps available at lower prices by showing users personalised advertising and selling their data to third parties, tracking can also have highly disproportionate effects on the lives of individuals and society as a whole. For example, tracking has significant effects on the rights to privacy and data protection, but also on other fundamental rights, such as the right to non-discrimination (e.g. when data from mobile tracking is used in AI systems, such as targeted ads for job offers) or the right to free and fair elections (e.g. when political microtargeting is used, as in the Brexit vote or the Trump election).

This thesis develops and applies techno-legal methods to study choice over app tracking at four levels: the impact of the GDPR (Chapter 4), consent to tracking in apps (Chapter 5), differences between Android and iOS (Chapters 6), and the impact of Apple's App Tracking Transparency (ATT) framework (Chapter 7). While many previous studies looked at data protection and privacy in apps, few studies analysed tracking over time, took a compliance angle, or looked at iOS apps at scale. Throughout our analysis of apps, we find compliance problems within apps as regards key aspects of US, EU and UK data protection and privacy law, particularly the need to seek consent before tracking. For instance, while user consent is usually required prior to tracking in the EU and UK (under the ePrivacy Directive), our empirical findings suggest that tracking takes place widely and usually without users' awareness or explicit agreement.

This thesis contributes 1) a scalable downloading and analysis framework for iOS and Android privacy and compliance analysis (PlatformControl), 2) an improved understanding of the legal requirements and empirical facts regarding app tracking, 3) a comprehensive database of the relations between companies in the app ecosystem (X-Ray 2020), and 4) an Android app to support the easy and independent analysis of apps' privacy practices (TrackerControl).

Contents

1 Introduction

Over the past decade, smartphones have revolutionised how we use technology. Individuals can now find and install the right app for millions of use cases in a matter of seconds. Smartphones have not only rebalanced how we communicate with our peers, but also how we work, love and live. Many even share their most intimate and vulnerable moments with them. This, in turn, makes smartphones the source of some of the most pervasive and valuable data and has helped create the modern surveillance economy [1–3]. This app data is currently being amassed by both a handful of tech companies and many opaque, smaller data brokers in order to derive vast profits but with yet unmitigated societal externalities [1, 4].

The harms that arise from the ongoing mass-scale collection of users' behavioural data – known as *tracking* – are varied and highly individual. They include threats to individuals as much as threats to society as a whole. Individuals often have no real choice over their data (e.g. in the absence of consent banners, or when dark patterns are used [5, 6]). When they are given a choice, this is often ineffective (e.g. when consent banners assume user consent regardless of a user's choice [7]). Further, individuals are often not fully aware of the consequences of their choices over data [8, 9]. Meanwhile, large-scale user tracking by apps is key to many lucrative black-box technologies, such as recommender systems, online behavioural advertising and

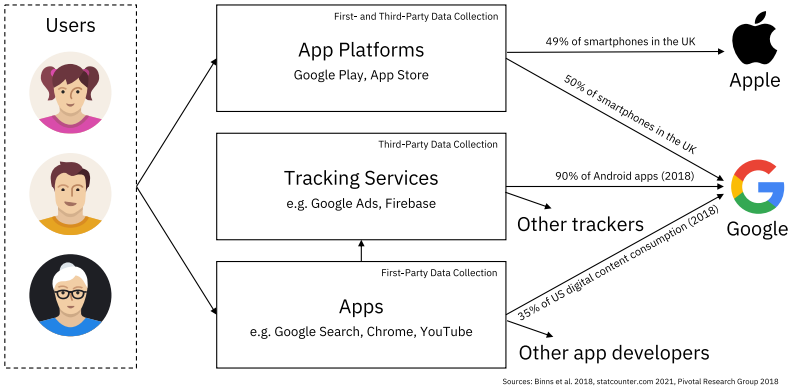


Figure 1.1: Schematic overview of data flows in the mobile tracking ecosystem. Google pursues vertical integration across the app data ecosystem, while Apple instead aims for horizontal integration.

microtargeting. These practices have been demonstrated to polarise online and offline discourse [10, 11], threaten the integrity of elections [12, 13], discriminate against disadvantaged groups [14, 15], and help design highly addictive and distracting technologies [16, 17]. More generally, many individuals feel like they have lost control over their personal data and privacy, leading to frustration and resignation regarding the design of fundamental digital technologies of the 21st century [8, 18], thereby discouraging democratic deliberation around those very technologies.

It is an open secret that those end-users who use smartphones often have limited choice and transparency about how apps treat their data [5, 19]. Despite this, we have surprisingly few empirical insights into the extent to which individuals have *some* choice. How many apps ask their users for consent to such tracking? Does it matter whether one chooses Android or iOS? Has the introduction of the General Data Protection Regulation (GDPR) in Europe in 2018 or of new privacy measures by Apple in 2021 changed the extent to which individuals have a choice over tracking?

Motivated by these observations, this DPhil thesis addresses the following three-part research question:

> *What choice do individuals have over app tracking and how can we measure this with techno-legal methods; to what extent have notable interventions (such as the introduction of the GDPR and Apple's App Tracking Transparency framework) changed the status quo; and what are the implications for current practice in technology and law?*

The first part of the research question, the *techno-legal analysis* of choice over app tracking, aims to understand the status quo better. There remain important gaps in our current understanding, especially as regards the differences between iOS and Android in terms of privacy protection and the implementation of consent in apps. Methodologically, this thesis aims to find answers to this part of the question by applying methodologies from computer science to the compliance domain around apps, and thereby create novel contributions.

The second part of the question will expand on the methods of the first part but will additionally consider the *longitudinal* dimension of app tracking and privacy, given that there have been a range of notable interventions to improve privacy protections – notably the GDPR introduced in May 2018 and Apple's App Tracking Transparency framework from April 2021. This analysis shall contribute to our understanding of what interventions are effective at regulating privacy problems in the app ecosystem, and what key challenges remain.

The last part of the research question will summarise and contextualise the results of this thesis, and derive concrete recommendations for the relevant decision-makers working in technology and law. Specifically, we identify six priorities to improve current practice in technology regulation.

By no means shall this dissertation put too much emphasis on individual choice, which has been championed as a magical cure for what is a deliberate lack of data protection and good faith in the surveillance economy (e.g. the DAA's AdChoices [20], IAB's Transparency and Consent Framework [21], W3C's Do Not Track [22], Global Privacy Control [23], Apple's App Tracking Transparency [24]). To effect fundamental change to the current data ecosystem, an individual's choice will hardly make any difference [3, 25–27]. Rather, good insights and methods are needed to inform the ongoing debate and call out the industry's fig leaves, motivate the emergence of new and improved software design and policy, and help the authorities enforce the existing legal framework at scale across millions of apps. The development of such *insights* and *methods* is the primary aim of this dissertation.

While there is extensive debate about the exact definition of tracking and what kinds are acceptable (e.g. considering crash reporting and certain forms of app analytics as acceptable, but not most personalised advertising technologies), this dissertation considers those debates a distraction and intentionally takes a broad definition of classifying all large-scale collection of users' behavioural data, especially if not explicitly asked for by end-users, as tracking. State-of-the-art machine learning techniques make it possible to derive highly accurate predictions (but sometimes also highly inaccurate ones) from whatever kinds of data are fed into the algorithm. More data is usually better. Instead of getting lost in the debate of what tracking is acceptable (an agenda that the industry keeps pushing and that is much debated in the legal and policy discussion) or of what kinds of data are collected by apps (a debate that technical scholars often focus on since this is relatively easy to observe), this thesis instead takes the use of tracking in apps as a proxy for the relative power that actors have in contemporary digital technologies in forcing their own terms, often illegally, upon end-users.

The focus on tracking (and acceptability of such) in this thesis is primarily legal, but also one that is critical of immense power in tech that faces limited checks and balances. It is driven by an observation that there seems to be a widespread disregard for applicable legal norms. This is somewhat puzzling. Just like cars are expected to fulfil all relevant product safety laws, one might expect that the same would be true for fundamental digital technologies; this is not currently the case, as found in this and other research. When reporting about these data practices, this thesis will sometimes engage in normative judgements. While some researchers hold that research should be 'objective', such an assumption would be naïve for this kind of research. Research, unless extremely theoretical and detached from this world, is never free from researchers' biases. Indeed, the explanation and normative contextualisation of observations is especially important in the context of highly invasive tracking with substantial societal effects; daily disinformation about and disregard for the requirements of the GDPR (which does not require consent on every website, despite us being constantly subjected to 'consent' pop-ups); and a

research area with significant uncertainty, imbalance of power and limited insights into the underlying technical systems. If this thesis was confined to objective and measurable arguments, then there would not be much to say beyond what types of data apps collect. This would hardly be helpful. Instead, normative judgements are made with the greatest care in this thesis and are limited as much as possible to the Discussion and Conclusions sections, where the collected findings are summarised, contextualised and explained.

The following shall describe this DPhil research – its motivation, methodology, and contributions – in more detail.

1.1 Structure

We first review, in Chapter 2, the relevant literature and give an overview of the tracking ecosystem at large. This introduces the reader to the topic of this thesis and provides the necessary background for the rest of this dissertation. This review will be loosely guided by Lawrence Lessig's [28] four modalities of cyberspace to ensure a comprehensive review of all relevant aspects of the mobile tracking ecosystem.

Following up in Chapter 3, we introduce our app analysis method that is the foundation for the rest of this dissertation. Core aspects of this are the 1) app dataset of 2.2 million apps, including iOS and Android apps, and apps from before and after the introduction of the GDPR in May 2018, 2) the X-Ray 2020 database with detailed information on tracker companies, and 3) the PlatformControl analysis toolkit. We restrict this Chapter to a high-level overview of the analysis methods used in the later Chapters in order to give the reader a point of reference and guidance. We later provide more details on the methodology to answer sub-questions arising from the main research questions.

Next, we conduct four studies, one for each sub-question:

1. Law: How has the GDPR affected user choice over tracking? (Chapter 4)
2. Apps: What choice over tracking do apps give users? (Chapter 5)

3. Operating System: How can the choice between Android or iOS affect tracking? (Chapter 6)
4. Platforms: How have Apple's recent privacy measures affected user choice over tracking? (Chapter 7)

By triangulating user choice from different perspectives, we aim to give a holistic overview of the state of user choice over tracking in the mobile ecosystem, as well as generate rich findings for regulators, policymakers and the general public.

The first sub-question, on the impact of the GDPR on apps' data practices in Chapter 4, is relevant given that many see this law as the 'gold standard' of data protection and privacy legislation [29]. This Chapter illustrates the shortcomings of current EU/UK legal remedies against apps' invasive data practices through an empirical analysis of apps' data practices at scale.

The second sub-question, on consent in apps in Chapter 5, analyses whether apps themselves implement measures to give users choice over their data practices. If they did, one could argue that no further interventions for more user choice over data would be necessary. However, we find that many apps do not provide users with *any* choice over tracking. Many app developers continue to rely on invasive data practices and are unable to move away from them; others are not even aware of the compliance obligations that they face [30–32]. This underlines the need to look beyond individual app developers to effect meaningful change to current data practices, and towards more compliance.

The third sub-question, on mobile operating systems in Chapter 6, scrutinises how Android and iOS apps compare in terms of a range of important aspects that relate to privacy and compliance. This is motivated by the fact that there is increasing competition between the two platforms in terms of privacy. Apple even claims 'Privacy. That's iPhone.' in its marketing campaigns, but provides rather limited evidence and restricts independent research efforts into this topic through various means as we will discuss. The last large-scale study on privacy in iOS apps had been done in 2013 [33], which further underlines the need for

renewed study of iOS and the development of reproducible research methodologies for continued analysis in the future. This strand of research reveals that iOS – at best – provides marginally better levels of privacy and compliance compared to Android and that they share many of the same problems.

The last sub-question, looking into Apple's App Tracking Transparency (ATT) framework from April 2021 in Chapter 7, analyses the extent to which app platforms can themselves act on privacy. This is important because the ATT sparked an unprecedented public clash between Apple and other large tech companies, notably Facebook, over opt-in popups to tracking. This is somewhat surprising, given that many forms of app tracking have needed consent since the amendment of the EU ePrivacy Directive in 2009. Specifically, this law requires consent for access and storage of information on an individual's terminal device under Article 5(3) (see the legal analysis in Chapter 5).

We close this dissertation by discussing the implications of the conducted studies in Section 8. As part of this, we derive a set of recommendations that emerge from the research in this work on how to improve the current mobile ecosystem as well as the regulation of digital technologies more generally. Given the complexity of the digital economy, it should be clear that there will be no simple 'fix' and this thesis does not attempt to put one forward.

1.2 Contributions

This DPhil research developed the first analysis framework to study app privacy and compliance at scale, over time, and across iOS and Android – which we call *PlatformControl*. The insights generated from this framework offer new perspectives into architectural and regulatory issues in the smartphone ecosystem regarding data protection and competition. The findings from this research have been submitted to EU, UK and German regulators and have been received with great interest. Crucially, the data and methods developed in this thesis have been made public at https://platformcontrol.org/ in order to motivate and enable follow-up

research. The publication of these methods is important since any quantitative study into the app ecosystems will be outdated by the time of publication. Besides the core DPhil research in large-scale app analysis studies, we also made some of our analysis tools available to everyday app users through the release of the *TrackerControl* app. This app uses the same database of characteristic tracker signatures that we used for our large-scale app ecosystem studies ('X-Ray 2020'), and has been downloaded more than 100,000 times. A supportive community of users has even translated the app into 21 languages.

1.3 Publications

The following lists all publications that have been (in part) the foundation for this DPhil dissertation:

- (Chapter 3.2.2) Konrad Kollnig and Nigel Shadbolt (2022). TrackerControl: Transparency and Choice around App Tracking. *Journal of Open Source Software, 7*(75), 4270. [34]
- (Chapter 4) Konrad Kollnig, Reuben Binns, Max Van Kleek, Ulrik Lyngs, Jun Zhao, Claudine Tinsman and Nigel Shadbolt (2021). Before and after GDPR: tracking in mobile apps. *Internet Policy Review, 10*(4). [35]
- (Chapter 5) Konrad Kollnig, Reuben Binns, Pierre Dewitte, Max Van Kleek, Ge Wang, Daniel Omeiza, Helena Webb and Nigel Shadbolt (2021): A Fait Accompli? An Empirical Study into the Absence of Consent to Third-Party Tracking in Android Apps. *SOUPS 2021.* [6] *(Winner of the Student Paper Award, Privacy Papers for Policymakers 2022, Future for Privacy Forum)*
- (Chapter 6) Konrad Kollnig, Anastasia Shuba, Reuben Binns, Max Van Kleek and Nigel Shadbolt (2022): Are iPhones Really Better for Privacy? A Comparative Study of iOS and Android Apps. *PETS 2022.* [36]
- (Chapter 7) Konrad Kollnig, Anastasia Shuba, Max Van Kleek, Reuben Binns, Nigel Shadbolt (2022). Goodbye Tracking? Impact of iOS App Tracking Transparency and Privacy Labels. *ACM FAccT 2022.* [37]

- (Chapter 8) Konrad Kollnig (2023). Lessons from the GDPR: Five priorities for effective IT regulation. *Ad Legendum, 2023*(2). [38]

While aspects of the work reported in this DPhil were carried out in collaboration with other researchers, in all cases, the research was led by the first author, including planning, managing and writing up.

2

Background: Mapping the Mobile Tracking Ecosystem

Smartphones are among the most ubiquitous pieces of technology. They bring great benefits – such as increased connectivity and productivity – but also harms, including to our data and attention. An important reason for these benefits and harms is that this technology regulates human behaviour in new ways.

Lawrence Lessig popularised the idea that the design of software can have similar effects as enacting laws, calling it *'code is law'* [28]. He identified four modalities that regulate human behaviour: norms, architecture, markets, and law. Norms mean the complex net of (often unwritten) rules within society. Nature provides the architectural constraints of human behaviour. The modalities in the digital world – which Lessig calls *'cyberspace'* – are similar, with the exception that code defines its architecture. Lessig further argues that code makes the digital world heavily regulatable. Code can be changed easily. In this, Lessig identifies a unique opportunity for society to write the rules of the digital world. The malleability of code also means that it may be written against us, with limited opportunity for change once standards are established. Lessig concludes that, if we want to preserve the freedoms of liberal democracy in the 21st century, we must take the chance to regulate code before it is too late. Otherwise, a few private powerful actors might

take up this regulatory space – something that we arguably see in the current world where a few platforms write the foundational laws for the digital world in code.

Similarly, Yochai Benkler identifies many unique opportunities in the digital world. In his book *The Wealth of Networks* [39], he argues that the high degree of division of labour on the Internet, paired with new non-monetary incentives, challenges classical economic thinking, and can create vast efficiencies. He observes that the corporate behemoths of the past century try to rewrite the rules of the digital world so that their outdated – often proprietary – business models prevail. This threatens the enormous efficiencies of the digital world.

The following Sections use the framework by Lawrence Lessig [28], and describe the mobile tracking ecosystem along the four modalities 1) norms, 2) code, 3) market, and 4) law. We choose this framework to consider the different forces at stake in mobile tracking, and to ultimately understand the levers to enact change to the ecosystem so that user data might be protected better.

Among other aspects, we explore *informational self-determination* as a driver of individuals' decisions in using technology and how individuals often struggle in practice to exert agency over their data in the Section on 'Norms'. The Section on 'Code' explores what principal parts make up the technological infrastructure of the tracking ecosystem; we also review methods for app analysis that directly inspect the app code (*static analysis*) or observe app behaviour at run-time (*dynamic analysis*). In the Section on 'Market', we explore the incentives of the market participants in the tracking market as well as the current distribution of power within this ecosystem. As part of this, we review the literature on platform studies and work on app platforms – i.e. on Google and Apple's app ecosystems. Lastly, in the Section on 'Law', we review the relevant legal framework in the EU and UK, particularly data protection law (the GDPR and the ePrivacy Directive), competition law in the context of digital platforms, and targeted platform regulation.

2.1 Norms and Struggles around Informational Self-Determination

2.1.1 Informational Self-Determination and Limits

An important non-monetary incentive in the digital world is the management of data about oneself. Individuals strive for self-determination, including control over the flows of information concerning them. From this, there has emerged the *right to informational self-determination*, first formalised in law by the German Constitutional Court in 1983 [40]. Today, this right is known as *the right to data protection*, and protected by the General Data Protection Regulation (GDPR) in the UK and EU.

Informational self-determination falls into the wider objectives of liberal democracies to embrace individual choice as the main source of power. Individual choice drives both politics and markets, through elections and purchases, respectively [41]. As an approach to privacy protection, consent is associated with the regime of *notice & choice* [25], which similarly relies on individual choice. For many data-processing activities, companies that want to process data from an individual must

1. Adequately inform the individual (*Notice*), and
2. Obtain consent from the individual (*Choice*).

These two fundamental requirements are often implemented in software through the provision of a privacy policy, accompanied by consent options for the end-user.

The limitations of the notice & choice paradigm have been explored in a range of scholarship. Regarding 'notice', it has been documented that most people do not read privacy policies, and that when they try to, have difficulties understanding them [42] and do not have enough time to read every such policy [9].

Regarding 'choice', evidence suggests that many individuals struggle with privacy decisions in practice [8, 43]. The mismatch between stated and observed privacy preferences is known as the 'privacy paradox' [44], although this so-called 'paradox'

may be best explained by structural forces that prevent alignment between values and behaviour [45, 46]. Individuals often have no real choice but to accept certain data processing because some digital services – such as Facebook or Google – have become indispensable [19]. Even when offered genuine choice, individuals face ubiquitous tracking [47], are tricked into consent [5], and do not get adequate compensation in exchange for their data [48]. Because of the limits to individual privacy management, various scholars argue that the regime of notice & choice does not provide *meaningful* ways for individuals to manage their privacy [25–27]. However, this past research does not mean that individuals do not care about their privacy, but, instead, that they are often faced with choice architectures that exploit their known psychological biases [5, 7].

There is a deeper challenge to the individualistic management of personal data. Individual choice over data has profound limits in the age of big data. Data collection and processing at scale reduces individuals to statistical predictions, with limited judicial safeguards [49], and can lead to new forms of discrimination [14]. The choice of an individual becomes increasingly meaningless [25, 27], and personal data increasingly social [50, 51]. Indeed, privacy research has long acknowledged social norms and interactions in shaping privacy expectations, as does Helen Nissenbaum's influential theory of *contextual integrity* [52]. This underlines how challenging it is to preserve data protection rights in the digital world, and the need to look beyond the individual.

2.1.2 Conflicts around Data Protection in Practice

The conflict between societal norms and data protection becomes most apparent when considering the use of social media. While it is known that many social media providers monetise user data and have previously suffered from immense data leaks (e.g. the Cambridge Analytica scandal on Facebook), individuals often have limited choice but to use such digital services to interact with their friends, in part due to peer pressure, in part due to lack of good alternatives [19].

While most individuals do care about how their data is used [44, 53], data protection is not their only interest and they often give in to short-term gratification [54]. The use of data to show targeted advertising even allows developers to offer apps at lower prices, often for free. An outright ban on all advertising in apps might not be the most adequate way forward, since it would prevent many individuals from accessing valuable apps and potentially deepen the digital divide [55]. The adequate use of data by apps therefore depends as much on the privacy preferences of individuals as well as their financial means.

Famously, Shoshana Zuboff has coined the term 'Surveillance Capitalism' to encapsulate the business model of online services [3]. Due to strong economic incentives, these online services are ever more driven towards data collection, while disregarding the privacy interests of individuals and the implications for democracy more generally.

2.1.3 Heightened Privacy Expectations of Individuals

The public is becoming increasingly interested in how companies and authorities treat their data [56, 57], especially since the Edward Snowden leaks in 2013 and the Facebook-Cambridge Analytica revelations in 2018. The Snowden leaks were perceived as a key moment by many individuals, including parliamentarians. As a consequence, the preamble of the GDPR, the data protection law of the EU from 2016 (see more in Section 2.4), explicitly states that it seeks to tackle the 'widespread public perception that there are significant risks to the protection of natural persons, in particular concerning online activity'. Tracking is one such risk [4, 58, 59], especially since US intelligence agencies can access the data troves of Google and other tech companies. The risks posed by US intelligence have made the European Court of Justice restrict the sending of personal data to the US, as part of its 2020 *Schrems II* judgement [60] (more discussion in Section 8.3.1). Whether the proposed EU-U.S. Data Privacy Framework from 2022 will address the problems highlighted in *Schrems II* remains to be seen.

As a result of the heightened privacy expectations of the public, gatekeeper companies have been forced to rethink their data practices. Prominent recent examples are the introduction of the App Tracking Transparency framework on iOS (blocking access to unique user identifiers without user consent), the planned ban of third-party cookies from the Google Chrome browser (preventing websites from saving unique identifiers in cookies to track users across websites), and Google's introduction of a user opt-out from sharing personal identifiers with apps on Android from the end of 2021. This increased competition around privacy is a choice of consumers, and shows that consumers can gather to demand better terms around personal data – despite the privacy paradox that mainly arises when individuals do not have a genuine choice. This difficulty in choice, then, is an important foundation of data-driven business models of privacy actors.

2.2 Code and App Analysis Methods

The code to collect data about smartphone users comprises four main pieces: apps, libraries, app stores and operating systems. We first traverse these four different parts of the code behind the tracking ecosystem, and then focus on the methods to analyse the data practices of *apps*.

2.2.1 Apps and Libraries

Apps are the medium for many tracking technologies. Such tracking is usually implemented through third-party *libraries* that serve as plug-in components for app developers. These libraries, in turn, collect data about users during their app usage, and send this data to the servers of trackers. The main advantage of this approach is that it streamlines the processing and collecting of user data. However, the use of these libraries and backend servers also necessitates a further third party (as opposed to the developer implementing the services by themselves) and often processes user data in closed-source systems, with limited transparency for the other stakeholders.

2.2.2 App Stores

App stores serve as an important part of the tracking ecosystem. They are the first point of contact for users with apps that these users wish to install. App stores also provide some information about apps' privacy practices, including privacy policies, privacy nutrition labels, and app permissions. However, there exists no readily available information on these app stores about the tracking software used by apps, nor the ability to filter for apps that come without tracking or without internet access at all. That this is possible is demonstrated by the Aurora and F-Droid stores on Android; these can be installed manually alongside the Google Play Store, but are not allowed to be shipped through Google Play itself, due to Google's policies. App stores are also involved to some extent in the tracking of users. Both Apple ('App Analytics') and Google ('Play Console') collect detailed download statistics and make these accessible to developers, thereby providing basic analytics to developers even without the integration of analytics tracking libraries.

2.2.3 Operating Systems

Beyond the app stores, the operating systems are also involved in app tracking. On iOS, the system library SKAdNetwork facilitates the attribution of clicks on ads, without using user identifiers, thereby reducing the ability of trackers to build profiles about users. However, this 'privacy-preserving' approach by Apple also discloses information about users' ad clicks to Apple, which in turn could use this data to build profiles about users for its own advertising business. Indeed, the company claims in its privacy policy that it might use users' 'interactions with ads delivered by Apple's advertising platform' [61], which might include third-party ads that use the SKAdNetwork. Furthermore, the iOS operating system provides users with various privacy choices, including the ability to manage apps' background data transmissions, location access and use of tracking.

In the Google Play ecosystem, most of the OS-level tracking software is shipped through the Google Play Services, which are necessary to access the Google Play

Store. Similar to iOS, this part of the Android system allows users to limit the extent of tracking, though not to ban it completely. Furthermore, the Google Play Services facilitate data collection as part of Google Ads, Google Analytics and Google Firebase Analytics [62], with currently no options for end-users to disable this data processing. This implies that every Android phone with access to the Google Play Store comes with extensive tracking functionality. This raises concerns about whether end-users are fully informed and have given consent to this practice at the point of making their purchase decision [63]. The French data protection regulator CNIL found that this used to not be the case, and fined Google over its design of the Android ecosystem and the lack of transparency in 2019 [64].

Previous research studying the data sharing by the operating system itself found that Android and iOS each share data with Google and Apple at high frequency, with great detail, and with limited user choice [65]. This data collection goes beyond the examples mentioned in the previous paragraphs. It would be beyond this thesis to explore all the Android and iOS components that can be used for user tracking.

2.2.4 Analysing Data Use in Apps

To gather meaningful insights into data protection in apps, one needs adequate analysis tools. Previous research analysing apps' data practices through technical means tends to fall into two categories, dynamic and static analysis.

2.2.4.1 Dynamic Analysis

Dynamic analysis observes the run-time behaviour of an app, to gather evidence of sensitive data leaving the device. Early research focused on OS instrumentation, i.e. modifying Android [66] or iOS [33]. Enck et al. modified Android so that sensitive data flows through and off the smartphone could be monitored easily [66]. Agarwal and Hall modified iOS so that users were asked for consent to the usage of sensitive information by apps [33], before the introduction of run-time permissions by Apple in iOS 6.

With growing complexity of mobile operating systems, recent work has shifted to analysing network traffic [2, 67–72]. Ren et al. instrumented the VPN functionality of Android, iOS, and Windows Phone to expose leaks of personal data over the Internet [71]. Conducting a manual traffic analysis of 100 Google Play and 100 iOS apps, they found regular sharing of personal data in plain text, including device identifiers (47 iOS, 52 Google Play apps), user location (26 iOS, 14 Google Play apps), and user credentials (8 iOS, 7 Google Play apps). Van Kleek et al. used dynamic analysis to expose unexpected data flows to users and design better privacy indicators for smartphones [2]. Reyes et al. used dynamic analysis to assess the compliance of children's apps with COPPA [70], a US privacy law to protect children. Having found that 73% of studied children's apps transmit personal data over the Internet, they argued that none of these apps had obtained the required 'verifiable parental consent' because their automated testing tool could trigger these network calls, and a child could likely do so as well. Okoyomon et al. found widespread data transmissions in apps that were not disclosed in apps' privacy policies, and raised doubts about the efficacy of the notice & choice regime [73] (as discussed in the previous Section).

Dynamic analysis offers different advantages. It is relatively simple to do, largely device-independent, and can be used to monitor what data sharing actually takes place. It also comes with limitations. The information gathered might be incomplete if not all code paths within the app involving potential data disclosures are run when the app is being analysed. Network-based dynamic analysis may wrongly attribute system-level communications to a studied app, e.g. an Android device synchronising the Google Calendar in the background, or conducting a network connectivity check with Google servers. Network-based dynamic analysis is nonetheless a versatile, reliable and practical approach.

2.2.4.2 Static Analysis

Static analysis dissects apps without execution. Usually, apps are decompiled, and the obtained program code is analysed [74, 75]. The key benefit of static analysis is

that it can analyse apps quickly, allowing it to scale to millions of apps [4, 59, 76–78].

Egele et al. developed an iOS decompiler and analysed 1,407 iOS apps. They found that 55% of those apps included third-party tracking libraries [75]. Viennot et al. analysed more than 1 million apps from the Google Play Store, and monitored the changing characteristics of apps over time [76]. They found a widespread presence of third-party tracking libraries in apps (including Google Ads in 35.73% of apps, the Facebook SDK in 12.29%, and Google Analytics in 10.28%). Similarly, Binns et al. found by analysing nearly 1 million Google Play apps that about 90% may share data with Google, and 40% with Facebook [47]. There has been some recent research on the new privacy nutrition and data labels on the app stores [79–81], but there are also concerns about the accuracy of these labels (see Chapter 7).

Static analysis can involve substantial computational effort and – unlike dynamic analysis – does not allow the observation of real data flows because apps are never actually run. Programming techniques, such as the use of code obfuscation and native code, can pose further obstacles. This is especially true for iOS apps, which are often harder to decompile – compared to Android apps – and are encrypted by default [36]. There are further reasons that make app analysis on iOS more difficult [4, 58], which we explore in Section 6.1.1. As a result, prior to the present thesis, the last large-scale study of iOS apps' privacy practices was done in 2013 [33].

2.3 Market: Platform Power in Digital Technologies

To understand the market of app tracking, we discuss the main stakeholders and their incentives around tracking. The main market participants (besides smartphone users, which we discussed in the previous Section on 'Norms') are developers, trackers and app platforms.

2.3.1 Developers and Trackers

Developers and trackers work closely together. Developers integrate the software of trackers into their own software for a variety of purposes. The most widespread mobile tracking technologies are developed by leading tech companies, foremost Google and Facebook; at the same time, there is a wide range of smaller, less-known tracking companies that also compete for access to user data [4, 35].

The use of third-party trackers benefits app developers in several ways, notably by providing analytics to improve user retention, and by enabling the placement of personalised advertising within apps, which often translates into a vital source of revenue [30, 31]. However, it can also make app developers dependent on privacy-invasive data practices that involve the processing of large amounts of personal data [31, 82, 83], with little awareness from users and app developers [25, 58, 70, 84].

Despite their crucial role within the software development life cycle, placing the blame for implementing consent incorrectly on app developers might be misguided. Many lack legal expertise, depend on the use of tracking software, and face limited negotiation power in the design of tracker software, which is usually developed by large, multinational companies [30, 31, 85–87]. At the same time, failure to implement appropriate consent mechanisms in software impacts individuals' choice over data collection and their informational self-determination, and may expose vulnerable groups – such as children – to disproportionate data collection [30, 88]. This underlines the need for robust privacy guarantees in code.

It is an open question to what extent app developers are liable under data protection law for their use of tracking. While trackers have an obvious interest in shifting responsibility for data processing onto developers, trackers arguably have more control over the design of the tracking ecosystem than individual developers. These, in turn, believe the responsibility lies with the trackers to ensure adequate levels of data protection [85, 87]. In February 2022, the Belgian Data Protection Authority (DPA) made an important ruling within this context that found that the Interactive Advertising Bureau (IAB) Europe had violated EU data protection

law, by unfairly shifting responsibility for data protection compliance onto software developers and publishers. IAB Europe has developed the most widely adopted technology for the retrieval and propagation of user consent to tracking, called *Transparency & Consent Framework* (TCF). The IAB had not adequately – according to the DPA – considered the rights and fundamental freedoms of the individuals affected by the design of their software [89]. We will explore the implications of this more in Section 8.3.1. IAB Europe has appealed the decision.

While developers of apps and the publishers of apps do not always coincide, for simplicity, we refer to both as 'developers' in this dissertation.

2.3.2 App Platforms

The hands of app developers and individuals in affecting app privacy are often tied (as discussed earlier in this Chapter), being faced with the interests of large data-driven companies. Because of this, it is natural to look at the two dominant app platform providers, Google and Apple, and review the existing literature on platform studies within this context.

2.3.2.1 Platform Studies

Much of the previous work derives from media and communication studies, and focuses on social media companies [90]. For instance, Tarleton Gillespie defines *platforms* as

> *'sites and services that host public expression, store it on and serve it up from the cloud, organize access to it through search and recommendation, or install it onto mobile devices'* [91]

This emphasises how platforms act as intermediaries of online discourse. They are not as neutral as it may seem, given that they often moderate conversation online, e.g. through search and recommendation functionality [92]. Indeed, Gillespie argues – discussing the example of YouTube – that digital intermediaries often use the term 'platform' for their ends, by suggesting a neutral role in information dissemination [93]. They face limited responsibility for user-generated content

under current law, and may need more regulation [91]. In regulating platforms, Gillespie warns against the *'fallacy of displaced control'*. Platforms cannot be blamed for all misbehaviour of their users.

Natali Helberger imagines Facebook as a powerful state with great power over public opinion [94]. Platforms are becoming ever more important political actors in the real world. Beyond the many indirect ways in which platforms moderate public debate, she points out that tech companies – including Google and Uber – have actively tried to mobilise their users for political ends, such as against the reform of EU copyright law. She concludes that more work from academia and policy is needed because the *'sheer possibility of the abuse of this immense power for one's own political goals is in itself a threat to any functioning democracy'*.

José van Dijck offers a more general view on platforms, focusing less on social media and online discourse. She defines platforms as

> *'the providers of software, (sometimes) hardware, and services that help code social activities into a computation architecture'* [95]

Similar to Lessig, this emphasises how platforms can shape social activity, and that this is done through technology.

In his book *Platform Capitalism*, Nick Srnicek offers a critical view on platforms. He considers them a new business model and reincarnation of capitalism [96]. Because of this, he argues that Apple is not a platform. It *'is more akin to the 1990s Nike business model than to the 2010s Google business model.'* After all, Apple's revenue mainly stems from device sales, not the App Store. Unfortunately, this focus on profits neglects some of the central externalities and resources in the digital ecosystem, namely data and attention (which are highly related due to the centrality of ads for digital business models, but pose different risks).

There seems to be a consensus among scholars that platforms exert some power over their users, and that they may use this power to advance their own objectives, rather than solely the users'. Unfortunately, previous platform research put limited focus on app platforms, and how their design might impede data protection rights

and influence human lives. This is despite the known challenges to data protection and the immense societal influence arising from smartphone technology.

2.3.2.2 App Platforms

Outside of China, there are two main providers of app platforms: Apple and Google. These two companies govern their respective app ecosystems, but pursue different strategies with respect to revenue streams, and the freedoms and responsibilities they grant to app publishers and users. In terms of revenue streams, both platforms take a share of up to 30% of all direct revenues created from app sales and in-app purchases, but differ otherwise [97]. Apple profits from the sale of iOS devices and does not licence iOS to other device manufacturers. Google's strategy, in contrast, is geared towards the global distribution of Android and Google Play on handsets manufactured by others [98]. Android itself is open-source, but Original Equipment Manufacturers (OEMs) pay a license to distribute the standard Google apps (including the Google Play Store app). A more significant source of revenue for Google is advertising; the parent company of Google, Alphabet, is estimated to have generated $147bn (80%) of its 2020 revenue from advertising [99], with more than half the revenue stemming from mobile devices [100]. This advertising business greatly relies on the collection of data about users, including from mobile devices. More users mean more data, which, in turn, results in more lucrative ads and revenue. While ads often give users access to software for free, the tracking and real-time bidding infrastructure that lie behind them are also known as a threat to individual privacy and can infringe on users' data protection rights [4, 77, 101, 102].

As well as differences in revenue streams, the two platforms differ in their approach to the level of freedom granted to app publishers and users. The Google Play Store grants relative freedom. End-users can modify their devices rather easily, and install apps from sources other than Google Play. The underlying operating system, Android, follows an open-source strategy, which has arguably contributed to its success [97, 98]. The freedom available on Android is not entirely unbridled. The open-source approach does not extend to many Google services

on Android, including the Play Services, upon which most apps depend for push notifications and in-app purchases (IAPs), among other essential functionality. Further, Google exerts discretionary control over apps admitted to the official Google Play store, which includes bans on certain types of apps, such as ad blockers. While no explicit justification is needed, the ban on ad blockers is based on the claim that such apps may 'interfere with [...] other apps on the device' [103]. However, in general, Google's restrictions have been much more permissive than those exerted by Apple on the iOS App Store, which has a much more stringent set of restrictions (e.g. regarding user privacy) and regularly uses manual review to check for compliance, using criteria that are not always clear [82].

These differences in revenue streams and control over app distribution are often cited to explain the alleged differences in the efforts each platform has made to restrict personal data flows and protect user privacy. Of the two, Apple has arguably placed a larger emphasis on privacy, seeking to gain a competitive advantage by appealing to privacy-concerned consumers [104]. For instance, as early as 2011, Apple started to phase out all permanent device identifiers, in favour of a user-resettable Advertising Identifier (AdId) – also called Identifier for Advertisers (IDFA). At their 2019 developer conference, Apple announced a ban on *all* third-party tracking from children's apps, a particularly vulnerable group of app users. In response to vocal industry concerns, Apple later backtracked from this absolute ban, and now still allows *some* minimally invasive tracking in children's apps. And in 2021, starting with iOS 14.5, Apple requires developers to ask users for permission before accessing the AdId or engaging in advertising-related tracking, sparking a fierce public battle between Apple and Facebook over tracking controls in iOS 14.5 [105, 106]. Facebook, like many other mobile advertising companies, is concerned that most users will not agree to tracking if asked more clearly and explicitly [107]; iOS users could already opt-out from the use of AdID, but were not explicitly asked by every app. While Google has followed Apple's lead in restricting the use of permanent identifiers, it currently does not allow all Android users to prevent apps from accessing the AdId (although Google is now rolling out a new opt-out system).

2.3.2.3 Regulation by App Platforms

Greene and Shilton provided one of the few studies we found that systematically compared the influence and roles of app platforms on shaping user privacy in 2018 [82]. They found that Apple and Google intervene in very different ways. While Apple imposes strict rules on its app ecosystem, which generally translates to user privacy benefits for most users, Google's ecosystem was mainly characterised by the absence of intervention. Google's approach has mixed implications for user privacy; 'The "wild west" of Android development means that privacy solutions abound for skilled hobbyists but that baseline privacy measures for the masses are lacking' [82]. These authors concluded that an app platform might be able to move quicker than the relevant authorities, being itself a *'privacy regulator'*; given the power of app platforms, they argued that there needs to be greater transparency around platform governance and enforcement of privacy rules. Similarly, Van Hoboken and Ó Fathaigh argued in 2021 [108] that Google and Apple increasingly act as important regulators of data protection and privacy, but with limited regulation, oversight and accountability. To increase transparency, these authors argued for mandatory disclosures about the privacy-related activities of smartphone platforms – as a minimally invasive but realistic intervention.

Zhou et al. hint at more concrete ways in which differences between app platforms can impact software development. They analysed bug-fixing in open-source projects, from desktop, Android and iOS [109]. They found that iOS bugs were fixed three times faster than on Android and desktop. The study only considered 16 open-source projects on iOS, as opposed to 34 on desktop and 28 on Android. This low number of iOS projects might be due to the potential incompatibility of Apple's App Store with certain open-source licences, such as GPLv2. Indeed, there have been reported instances of open-source apps having been removed by Apple from the App Store due to this incompatibility [110]. This provides an example of how app platform rules might impede the development of (open-source) apps.

There is some research in economics that analyses the different incentive structures on iOS and Android. Li et al. pointed out in 2020 that *piracy* is a big concern for app developers, and analysed the incentives for duopoly platforms around implementing strong piracy protection measures [111]. Roma and Ragaglia suggested in 2016 that paid monetisation models might be more lucrative than free ones on iOS, but not so on Google Play [112]. Wen and Zhu analysed in 2019 how the threat of the introduction of new Google-developed apps might influence prices and innovation [113].

There is also research in human-computer interaction that gives insights into the responsibilities of app platforms. In 2013, Kelley et al. conducted a user study to compare different ways to disclose app privacy in an app store [114]. They found that the *'question of trusting the information was one most [participants] had never considered, and actually gave some participants pause as they realized for the first time that this information might not be accurate. Again, users believe this information is correct, is being verified, and will assume they misunderstand something before they would believe the displays are incorrect.'* Users place great trust in the app store providers, but hardly question this trust.

Wetherall et al. observed in 2011 that users often have limited access to low-level information, which is important for privacy decisions [115]. Because of this, they suggested that the mobile operating system should establish transparency around apps' data practices. Similarly, Shih et al. criticised the fact that *'mobile platforms lack support for fine-grained control over data collection'*, and called for regulatory action in the absence of action by app platforms in 2015 [116]. As we will discuss in Section 2.4.4, there remains limited targeted regulation of app platforms even today, though this seems to be changing slowly.

2.4 Law: The GDPR and Other Regimes

We now introduce the legal background for app tracking. Since tracking relies on data collection about individuals, the most relevant laws relate to data protection

and privacy. In the EU and UK, the main pieces of data protection law are the General Data Protection Regulation (GDPR) and the ePrivacy Directive. These laws establish clear rules when it comes to the processing of personal data and provide additional safeguards when it comes to information relating to children. We focus on the EU and UK as they have relatively stringent and specific rules on consent and third-party tracking. While similar rules exist in other jurisdictions (such as the COPPA in the US, which requires parental consent for tracking), recent regulatory actions and rich guidance issued by European regulators offer an ideal setting for large-scale analysis.

2.4.1 General Data Protection Regulation (GDPR)

The GDPR (General Data Protection Regulation) came into force in May 2018 to protect data relating to individuals ('personal data'). It replaced the 1995 Data Protection Directive (DPD) aiming to address 'new challenges for the protection of personal data' brought abount by '[r]apid technological developments and globalisation', as stated in the preamble of the GDPR. Like the DPD before it, the GDPR places obligations on organisations that process personal data. Those who decide the means and purposes of such processing are 'data controllers', who are required to have a lawful basis for processing (e.g. consent or legitimate interests) (Article 6), and follow principles of fairness, transparency, purpose limitation, data minimisation, accuracy, security and accountability (Article 5). Those who undertake processes on behalf and under the instruction of data controllers are 'data processors' and have a less extensive set of obligations. All companies based in the EU and UK, as well as companies monitoring the behaviour of, or offering goods and services to, individuals located in the EU and UK, fall within the territorial scope of application of the GDPR (Article 3 GDPR).

In the context of third-party tracking, the first party (e.g. the app developer) is likely a controller; the third parties may be processors (where they only process data on behalf of the first party, e.g. for app analytics), controllers in their own right (where they use the first-party data for their own purposes such as targeted

advertising, improving their machine learning models, etc.), or sometimes both at the same time. While some third parties may present themselves as mere processors in order to avoid the obligations of a controller, recent case law of the Court of Justice of the European Union (CJEU) affirms that the bar may indeed be low enough to qualify many third parties as controllers (or joint controllers where they jointly decide on the purposes and means of processing). This has been confirmed, for instance, in the *Fashion ID* case [117], finding that when a website embeds a Facebook 'Like' button, which facilitates third-party tracking, it is a joint controller with Facebook; and the *Wirtschaftsakademie* case [117], where the operator of a Facebook fan page operator was deemed a joint controller.

We will further explore in Chapter 4 how the changes under the GDPR might have affected the extent of tracking in mobile apps.

2.4.2 ePrivacy Directive

Another important and relevant element of the data protection regime is the 2009 ePrivacy Directive; this covers the privacy of electronic communications and includes rules on the use of cookies and related tracking technologies. Under Article 5(3) of the ePrivacy Directive, third-party tracking typically requires consent as it involves accessing or storing data that is not strictly necessary for delivering the app or service's functionality on a user's device [6]. The ePrivacy Directive sits alongside and complements data protection law; it constitutes a *lex specialis*, meaning that, when both the ePrivacy Directive and the GDPR apply in a given situation, the rules of the former will override the latter. This means that even if it might otherwise be lawful to process data in third-party tracking under the GDPR without consent (e.g. using an alternative lawful basis like legitimate interests), the ePrivacy Directive would still require consent. Despite the UK leaving the European Union, both the GDPR and the ePrivacy Directive remain unchanged on the domestic UK statute books (at the time of writing), in the form of the UK GDPR and the Privacy and Electronic Communications Regulations (PECR).

We will further explore in Chapter 5 how the ePrivacy Directive requires the retrieval of user consent to most implementations of user tracking.

2.4.3 Competition Law

Competition law has long been a cornerstone of regulating the misuse of corporate power and ensuring the functioning of markets. On this basis, the EU has previously fined Microsoft €497 million for their deep integration of the Windows Media Player into their Windows operating system in 2007 [118], and Google €4.34 billion for its design of their Android platform in 2018 [119]. While these fines might seem high, they make up only a relatively small fraction of overall revenues of such large tech companies; these cases also tend to be highly complex and take many years, while the platforms can move on quickly – with potentially only small changes to their business practices. *New Brandeisian* scholars argue that current competition law, particularly in the US, does not go far enough to tackle the monopolistic threats of the 21st century. Instead of focusing only on economic indicators, competition regulators should increasingly include non-economic metrics [120–123].

In the context of the third-party tracking ecosystem, several studies have explored the extent of market concentration in data collection. Binns et al. [47] found that the current market concentration of tracking in apps and websites might warrant more attention from EU regulators. Google was identified as the most prominent data collector from websites and Android apps. Follow-up work studied the decisions of US and EU competition authorities regarding consolidation among tracking companies [123], finding that, by taking a traditional competition law focus on revenues and defined markets, they often failed to address the data dimension to market power and the resulting potential for abuse of dominance. An important reason for this lack of intervention is that the competition authorities did not consider it their responsibility to assess these transactions. Instead, data-related issues traditionally lie with data protection authorities, which do not usually assess M&A transactions. Lynskey [124] argues that EU data protection law is better suited than competition law to oppose such 'data power', but that further clarification and

integration of both legal regimes is needed, and that the current competition law fails to tackle the threat of data power. She argues that current EU data protection law implicitly suggests special responsibilities for large data-collecting firms, but these responsibilities need further clarification. Since neither data protection nor competition authorities are traditionally tasked with opposing the power (over data) of large tech firms, Binns and Bietti propose that we need a new regulatory approach in the EU and US, 'one that engages in a pluralist analysis of economic and noneconomic concerns about concentrations of power and control over data' [123].

Besides potential monopolies around (app) data, there also exists a duopoly in app markets, which has seen limited regulation as we will discuss in the next Section.

2.4.4 App Platform Regulation

The centrality of app platforms – i.e. Apple's iOS and Google's Android ecosystem – makes them a target for effective privacy regulation, however such regulation is limited [108, 125]. The US Federal Trade Commission (FTC) established some baseline rules for app stores in 2013. They strongly recommended app platforms to require just-in-time consent for sensitive data access, to seek privacy policies from app developers, and to implement a system-wide opt-out mechanism for data collection [126]. Despite not being law, Google and Apple followed many of the recommendations, and have not seen further public recommendations from the FTC since.

In the EU and UK, there, too, exists limited targeted regulation of app stores. The Regulation on platform-to-business relations (P2BR) contains general provisions for online intermediaries, including app stores, but does little to enact better privacy protections [125]. Data protection laws, such as the GDPR and the ePrivacy Directive, arguably place the primary responsibility for data protection with the app developers, not usually with app platform providers – although this is subject to ongoing debate; this lack of data protection obligations within the entire software development process – not just deployment – has been widely criticised [127, 128].

While limited targeted regulation exists, app platforms face increasing scrutiny by courts and regulators. In the case *Epic Games v Apple* running since 2020, a US District Court judge largely found no monopolistic behaviour from Apple, but did identify some anti-competitive conduct in Apple's business practices. The judge ordered Apple to allow app developers to inform app users of alternative payment methods. Both Apple and Epic Games have appealed the ruling. In the EU, following a complaint from Spotify against Apple in 2019, the European Commission identified multiple anti-competitive aspects of Apple's ecosystem in a preliminary ruling – the case is, however, still ongoing. In January 2022, the Dutch competition authority demanded changes from Apple to its App Store policies [129].

The challenges in keeping up with the regulation of platforms have spurred a recent countermovement by lawmakers. In South Korea, parliament amended the Telecommunication Business Act to force app stores to allow alternative payment methods and reduce commissions [130]. In response, Apple lowered the share it takes from App Store revenues of small developers (making less than $1 million per year) from 30% to 15%. In the US, Congress is debating a new Open App Markets Act that aims to address common competition concerns about app stores and passed the Senate Judiciary Committee with a strong 20—2 bipartisan vote in February 2022. In the EU, lawmakers, in late 2022, enacted two new pieces of legislation that aim to improve the regulation of digital markets, the Digital Markets Act and the Digital Services Act. Any new legal requirement for app platforms will likely have implications worldwide, due to the nature of digital ecosystems. It remains to be seen what impact these efforts will have.

2.5 Conclusions

Technology regulates behaviour. Some of the most ubiquitous pieces of technology are smartphones. These can pose a wide array of harms, including to the right to data protection. This is why it is important to have the right tools at hand to analyse how and to what extent end-users make decisions over their smartphone

tracking. Unfortunately, few studies have looked at iOS, architectural, or regulatory questions. Yet, an understanding of these dimensions is essential to track the health of the app ecosystem, understand its deeper problems, and assess the functioning of applicable regulation. There also exists limited work specifically on app platforms – rather, iOS and Android are often regarded as a natural extension of scholarship on social media platforms; this does not do justice to the importance and centrality of these platforms.

Mobile tracking is a multi-faceted ecosystem with a range of stakeholders, which each have their own special interests. When it comes to user privacy, app platforms have been making some promising changes in recent times, and regulators have increased their efforts in holding platforms' data practices to account. However, in all of this, the will of the end-users is essential, whether this is through purchase decisions (e.g. buying an Android or an Apple phone), casting their vote in elections (thereby legitimating EU and other politicians in putting forward new regulation), or by choosing free online services. Because the ultimate power over data practices lies with the end-users, this dissertation will zoom in on them in detail, and take a user-centred view. If something is to change about the app ecosystem, this will be driven by the end-users of the system.

From the above analysis, it is clear that no single measure will improve data protection on the ground. For this reason, this dissertation will not attempt to put forward 'fixes' for the ecosystem at large. Instead, this dissertation will try to establish matters of fact, and develop analysis techniques that combine technical and regulatory state-of-the-art considerations. In this analysis, because the ultimate choice rests with consumers, we will try to establish the status quo around user choice over mobile tracking. Since the status quo will change constantly, it is ever more important to have the right tools in place for constant assessment of the ecosystem. The main focus of the analysis in this dissertation will be the EU and UK because these jurisdictions provide relatively mature regimes around the regulation of data.

3
Analysis Framework

In the previous Chapter, we reviewed relevant prior literature and also introduced fundamental concepts of the mobile tracking ecosystem. A key aspect that emerged was a continued lack of transparency and understanding around app tracking, in particular concerning user choice and apps' compliance.

In this Chapter, we give an overview of the key elements of the methodology that is used and developed throughout this thesis. These key elements are:

1) **The app dataset of ~2.3 million apps.** The app dataset shall serve as a foundation for the app research in this DPhil dissertation and beyond.
2) **The X-Ray 2020 database.** This database contains comprehensive information about the structure of app tracking companies, their jurisdiction as well as the domains that they use for tracking. It has been derived from analysing the app dataset.
3) **The high-level analysis method ('PlatformControl').** The analysis framework PlatformControl provides the first toolset that allows the analysis of app privacy and compliance across iOS and Android at scale.
4) **The TrackerControl app**. This Android app makes some of the analysis methods from this DPhil research (particularly the X-Ray 2020 database)

	2017	2020
Android	959,426	1,066,400
iOS	-	294,917

Table 3.1: Number of apps in our app dataset, as well as the year of download. A few apps were already downloaded in December 2019, but are assigned to 2020, for simplicity.

available to non-expert Android users. This, in turn, allows individuals to collect real-time evidence of app tracking easily.

Wherever possible, the data and methods from this research have been made available at https://platformcontrol.org/. This is especially important since any findings from this research will be quickly out-of-date, due to the fast-changing nature of the mobile app ecosystem.

3.1 Data Collection

3.1.1 App Download and Dataset

This Section details our process for selecting and downloading apps from the Google Play and Apple App stores. This is a necessary precondition for any follow-up investigation in this thesis. To understand problems with app privacy and compliance at large, it is therefore pertinent to have a large dataset of apps as a foundation. If one seeks to study a specific subset of apps (e.g. children's apps), then a large dataset of apps would contain a sufficiently large and rather representative sample of apps from such a subset. See Table 3.1 for an overview of the app dataset used for this thesis.

App selection. To select apps, we fed the auto-complete search functionality of the Google Play and Apple App stores with alphanumeric strings of up to three characters to identify popular search terms, similar to previous literature [4, 76]. Searching for these terms on the app stores then allowed us to identify large numbers of apps and collect relevant meta information (including title, release date, and time of last update). We restricted our analysis to apps available in the

UK region for both app stores, on the basis that such apps must comply with the General Data Protection Regulation (GDPR). Despite the UK's withdrawal from the EU, the GDPR remained applicable in the UK at the time of our study, since it had already been adopted into national law. In total, we identified and retrieved the store entries of 1,459,422 Android and 696,039 iOS apps in December 2019. The number of identified apps on iOS is smaller, likely because the autocomplete functionality of the App Store returned fewer results and because the App Store contains fewer apps overall.

App download. Having identified a large number of apps, we then downloaded a subset of these apps. Due to storage constraints, we could not download all of the apps. iOS apps usually have a significantly larger size than Android ones, because they are already compiled into the different binary formats necessary for iOS devices (often containing both `arm32` and `arm64` instructions), while Android apps are shipped in an intermediary program language (Dalvik code). Our download methodology expands on the App X-Ray project, which is open-source [131] and had previously enabled the analysis of ~1 million Android apps in 2018 [4]. Adding to this project, we have 1) implemented a scalable download method for the Apple App Store, and 2) restored compatibility with the latest API changes of the Google Play Store to enable the download of Android apps at scale.

The X-Ray project used the existing Python library `gplaycli` [132] to download Android apps from the Google Play Store. For the Apple App Store, we used the automation tool `AutoHotkey` [133] to interact directly with Apple iTunes, through its Component Object Model (COM) interface. For each identified iOS app, a purpose-built `AutoHotkey` script opened the app's download page in the Windows version of iTunes and clicked the *Download* button, so as to download the app, similar to Orikogbo et al. [134]. In total, we downloaded 1,066,400 Android apps and 294,917 iOS apps over 2.5 months between December 2019 and February 2020. This is before the introduction of Apple's new opt-in mechanism for tracking in 2021. Our dataset therefore reflects privacy in the app ecosystem shortly before this policy change.

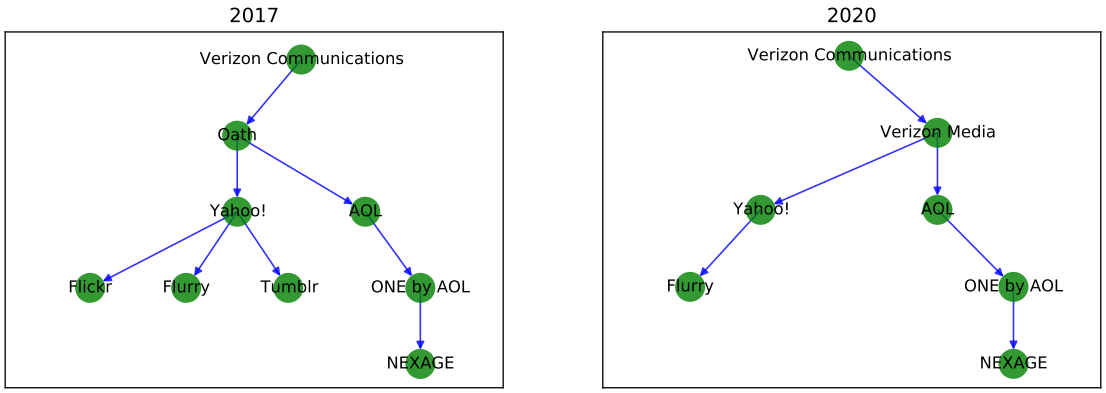


Figure 3.1: Company structure of Verizon's tracking business in 2017 and 2020, as an example of the diverse and changing nature of the tracking ecosystem. Only leaf companies present in at least 0.1

To protect against unexpected interruptions of the downloading process, we downloaded apps month by month with those apps last updated or released first. Unfortunately, due to Covid-19 and the subsequent closure of the Oxford Computer Science Department, our iOS download had to stop for apps when starting to process apps last updated in 2017. However, it was possible to continue the downloading of Android apps, and this interruption did not turn out to be a problem for our comparative analysis of the two app ecosystems in Chapter 6. In this Chapter, we assumed apps released or updated in 2018 or later likely cover the majority of apps currently in use, thereby sidestepping the problem of interruption. From this period, the number of downloaded iOS apps (n = 285,680) and Android apps (n = 283,065) was similar.

In addition to these newly downloaded apps in 2019/2020, we also used the previously downloaded 959k apps by Binns et al. [4] for our research. Specifically, we use these old apps by Binns et al. [4] in Chapter 4 to understand tracking in apps before and after the introduction of the GDPR.

3.1.2 X-Ray 2020 Tracker Database

When analysing tracking in apps, many pieces of previous research tended to focus on the quantitative presence of certain *tracking technologies* (e.g. how many apps contain 'Google Analytics'). While this is useful to gain insights into developer preferences

of certain tracking services, this approach struggles to characterise the reach of certain tracking *companies* fully. Insights into company relationships are especially important for questions that relate to competition and market concentration, which are receiving ever more attention from regulators and policymakers. At the same time, the tracker ecosystem is made up of a large and diverse set of tracker companies, some of which belong to or get acquired by other tracker companies [123]. For instance, Verizon Communications sold its subsidiaries Flickr and Tumblr, and restructured its online advertising business, see Figure 3.1.

To understand these diffuse company relations, a method for resolving tracking technologies to particular companies and the relationships between them is required. Binns et al. [47] previously created a database of known tracker companies and their company hierarchies in 2017, based on the analysis of 5,000 Android apps. For this thesis, we created *two* new and separate tracker databases, one for 2017 and one for 2020 apps, based on the previous database. For the 2017 database, we extended the existing database with those tracker hosts and libraries additionally found from our overall analysis in this thesis (as a result of the Chapters that follow), following the same protocols as the previous study by Binns et al. [47]. For each new tracker host or library, we checked to what company it belongs, what parent companies this tracker company has (using WHOIS registration records, Wikipedia, Google, Crunchbase, OpenCorporates, and other public company information), and in what jurisdictions these companies are based. We carefully included only those corporate relations that were already formed by the end of 2017 in the 2017 database. To create the 2020 database, we revisited every company in our 2017 database, and checked whether its ownership had changed. We used the same protocols as for the 2017 database to identify what companies are ultimately behind tracking.

Our systematic analysis of tracker libraries and hosts identified 24.4% additional companies (from 578 to 719 companies), comparing our 2017 database to the original database by Binns et al. [47]. Our 2020 database is slightly larger, and contains 754 companies, since it contains additional company transactions that have taken

place since 2017. We call the resulting dataset `X-Ray 2020`, and share it with the research community for follow-up studies.

3.2 Data Analysis

3.2.1 PlatformControl: Analysis at Scale and Across Platforms

This Section briefly introduces our analysis methods. We will develop these methods in more detail as we proceed in this dissertation, including the challenges faced and solutions found – in particular in Chapter 6 on comparing Android and iOS privacy and compliance. First, the analysis methods developed in this dissertation are capable of downloading iOS and Android apps at great scale, as discussed earlier in this Chapter in Section 3.1.1. This is not new per se, but had not been documented in the public domain in terms of ready-to-use code for the case of iOS or a unified toolset that targets both Android and iOS. As a result, few previous studies had focused on iOS – the last large-scale study of iOS app privacy had been conducted in 2013 [33]. Regarding the actual analysis method, we can derive a range of information from the analysed apps, for both Android and iOS apps. To achieve this, we combine *static* and *dynamic* analysis (as introduced in Section 2.2.4).

In the static analysis, we look at information about apps' privacy practices that can be derived directly from the app code, without the need for execution on a real smartphone. The information derived from this analysis includes the tracking technologies present in the app code, the types of personal data that apps can potentially access, and also the data-minimising configuration of tracking libraries. This static analysis allows us to study, at scale, the privacy practices of mobile apps (particularly in Chapter 6) and also the market structure of the tracking ecosystem at large (particularly in Chapter 4).

In the dynamic analysis, we run each app on a real smartphone and monitor apps' network traffic. This allows us to understand what companies receive personal data and what types of personal data (see Chapter 6). This also gives insights

into potential data sharing without user consent, which we study by opening each app but not interacting further with it (see Chapter 5).

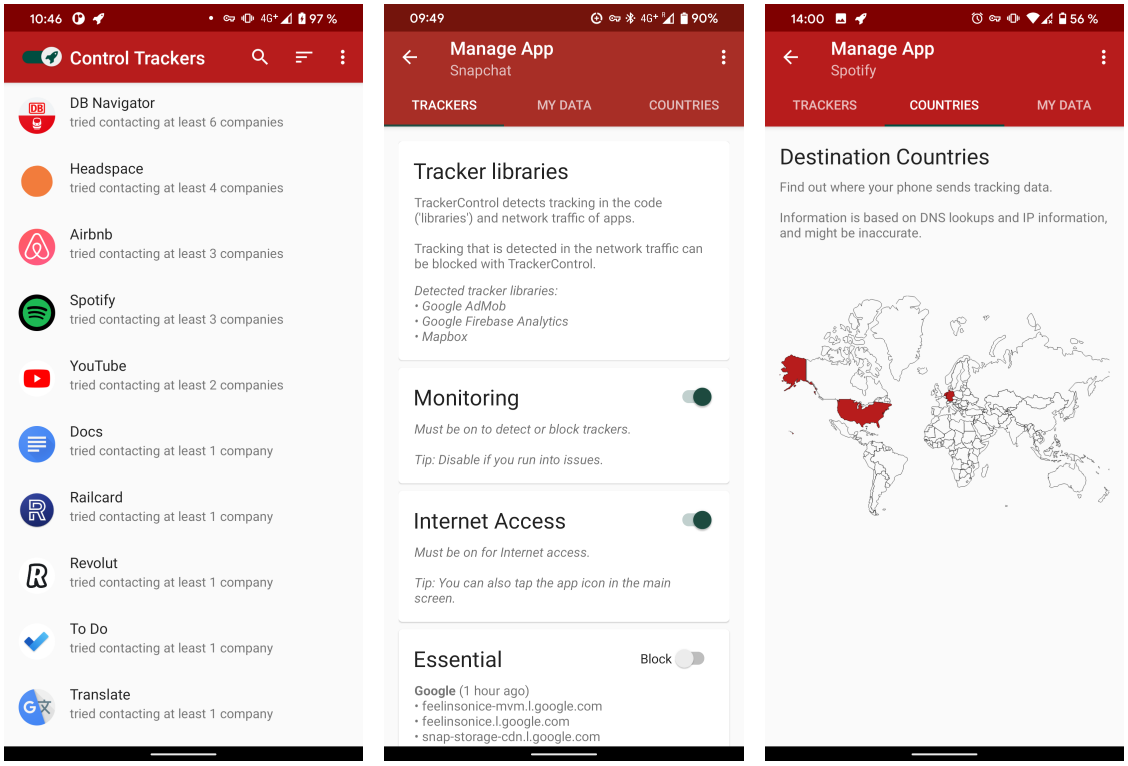
Our methods provide two key novelties: 1) Relatively worry-free, scalable analysis of iOS apps. Previous research has relied on app decryption, which is in a legal grey area; our iOS analysis method does not rely on decryption (see Section 6.1.1 for a detailed account) and provides similar performance as previous methods. 2) A focus on compliance within apps. This has rarely been addressed in previous research. As a result, we have been active in consultations with regulators and NGOs on the evolution of tech regulation as well as the application of existing tech regulation.

We make all our app download and analysis methods available online at https://platformcontrol.org/. This includes the first-of-its-kind database of signatures of tracker libraries on iOS (see Chapter 6), as well as the first-of-its-kind framework to grant permissions to iOS apps from the command line (this is important to analyse the sending of personal data over the Internet in an automated fashion). In the end, we hope that the publication of our methods will help motivate continued analysis of privacy and compliance in app ecosystems.

3.2.2 TrackerControl: Easy-To-Use App Analysis

This thesis develops a broad range of new analysis methods, for both Android and iOS. Most of these analysis methods require a real smartphone with elevated access (i.e. jailbreak on iOS and rooting on Android); this can, however, weaken the security guarantees of smartphones used in day-to-day life. Moreover, app analysis in research tends to be confined to lab settings, despite the best efforts of researchers.

To make our tracking analysis tools available to a wide audience and let individuals study their privacy in real-world situations, we developed the Android app TrackerControl (TC). This app provides users with real-time evidence of app tracking. TC analyses the network traffic of other apps by establishing a local VPN on the Android phone and matching all observed network traffic against a database of known tracking domains. This allows the generation of factual evidence



(a) Main screen: TrackerControl is enabled by clicking one button (at the top). If enabled, users can easily see how many companies apps share data with.

(b) For each app, users can see what companies and domains apps contact, block tracking by purpose (e.g. 'Advertising' or 'Analytics'), and disable Internet access. Additionally, users can inspect what tracker libraries are integrated into each app.

(c) Users are provided with further information, including the destination countries of their data and information about their GDPR rights (not shown).

Figure 3.2: TrackerControl aims to enable users to inspect easily what companies their apps share data with, and for what purposes.

of what companies apps share data with, and can support research (both academic and non-academic) on app privacy.

The tracking database behind TC is a unique feature of the app. The core of this database is the `X-Ray 2020` database which is the product of significant research efforts over the past years [2, 4, 6, 36]. This database has been introduced earlier above in Section 3.1.2. The `X-Ray 2020` is complemented by the Disconnect.me database which is the foundation for tracker blocking in Mozilla Firefox on the web. We further integrate the commonly used StevenBlack hostlist for tracking in apps, as a fallback. Overall, these databases provide information on 1) the *companies* behind tracking on the web and in apps, 2) the *countries* in which

these companies are based, and 3) the *purposes* for which tracking is conducted (e.g. analytics or ads). The visualisation of tracking inside TC loosely follows the work by Van Kleek et al. [2, 135].

The core of TC builds on the NetGuard app, which is in active use by millions of users worldwide [136]. The high maturity of NetGuard ensures the reliability of the tracker analysis while minimising battery impact and supporting the long-term maintainability of TC. In addition to providing insights into app tracking, users of TC can also block unwanted network transmissions, which has contributed to building a vibrant community of tens of thousands of users. This community has helped make TC available in 22 languages.

TC can also detect what tracker libraries are integrated into apps on a user's phone (i.e. static analysis). The foundation of this is the Exodus Privacy tracker library [137].

Beyond giving back to and connecting with the interested privacy community, TC was used for the research in this thesis in Chapter 5. TC has also facilitated and inspired other academic research at the intersection of policy and privacy technology [6, 138]. It has been used by the Finnish innovation fund Sitra for its 'Digipower investigation' into apps' data practices. As part of this study, leading Finnish politicians and journalists used TC to analyse the practices of Android apps. The results of this investigation were presented at the Finnish and European Parliament in 2022. TC is being considered to be added to the investigation repertoire of the French data protection regulator, and is used for citizen science on app privacy as part of the EU-funded CSI-COP project. Together with Dr Jun Zhao and other colleagues, we are currently developing a version of TrackerControl aimed at children, called the KOALA Hero app.

4

Tracking in Apps after the GDPR

To give citizens 'better control over how personal data is handled by companies and public administrations' [139], the EU updated its data protection regime with the General Data Protection Regulation (GDPR), brought into force in 2018. This law seeks to address, among other aspects, the risks posed by the widespread collection of personal data collection in apps, on the web, and in other digital contexts, by imposing specific requirements in the context of personal data processing. However, limited empirical evidence exists thus far regarding the effect the GDPR has had on the actual act of third-party tracking in smartphone apps.

In this Chapter, we examine the Android mobile app ecosystem, which remains the largest smartphone app ecosystem. We compare nearly two million Android apps from the UK app store, from before and after the introduction of the GDPR in 2018, to study how the tracking ecosystem has changed. Our data was collected when the UK was still bound to EU law – during the transition period of the EU-UK Withdrawal Agreement. Specifically, we examine the following three research questions:

1. How has the distribution of third-party trackers across apps on the Google Play Store changed?

2. How have the organisations doing the tracking themselves changed, in particular in terms of ownership and jurisdiction of operation?

3. How has the market concentration in third-party tracking changed?

These questions aim to understand, at a macro scale, whether the GDPR has thus far had a measurable and material impact on the tracking operations of smartphone data aggregators.

Our analysis suggests that there has been limited change in the presence of third-party tracking in apps, limited changes in ownership and jurisdiction of tracking companies, and that the concentration of tracking capabilities among a few large *gatekeeper* companies persists. However, significant change might be imminent, due to recent changes by gatekeeper companies.

Contributions. This Chapter makes important contributions to our understanding of the impacts of the GDPR. We provide large-scale quantitative evidence of how this law has affected an invasive and widespread data practice: tracking in apps. We provide new insights into competition within the tracking data market by applying the metrics by Binns et al. [47] at scale in 2m Android apps (previously 5,000 apps). We share all our code and data from this research at `https://osf.io/35xps/`.

Structure. The rest of this Chapter is organised as follows. We first explore why the GDPR might have affected the tracking practices in apps in Section 4.1. We introduce our methodology in Section 4.2 and our results in Section 4.3. We discuss our results in Section 4.4, the limitations of our approach in Section 4.5 and our conclusions in Section 4.6.

4.1 Implications of the GDPR for Tracking

We already introduced the key principles of the GDPR in Section 2.4. Based on this, we now explore further how the changes under the GDPR may have affected third-party tracking in apps, and thereby motivate the subsequent analysis of apps' data practices before and after the GDPR.

4.1.1 Key Changes under the GDPR

Previous data protection law was conceptually and formally very similar to the GDPR, and therefore the legal status and obligations of third-party trackers have not changed substantially [4, 40, 140]. However, several changes introduced by the GDPR could be expected to make a difference to the compliance efforts of third-party trackers on the ground. In the context of the compliance practices of third-party tracking, three categories of change are particularly pertinent: 1) stricter data protection standards, 2) new governance and accountability obligations, and 3) improved enforcement mechanisms.

Stricter data protection standards. The GDPR sets a higher bar for consent to data processing than the DPD (Articles 2 and 7 DPD; Article 7 GDPR). Under the GDPR, consent needs to be freely given, affirmative, specific, unambiguous, and informed. In the context of third-party tracking, these new consent standards have had the effect of enhancing the existing consent requirements under the previously introduced 2009 ePrivacy Directive (see Section 2.4). Specifically, the ePrivacy Directive requires user consent, according to the improved consent standards of the GDPR, for storing and accessing information on a user's device – a prerequisite for most forms of third-party tracking. Moreover, third-party trackers may now struggle to demonstrate their compliance with this consent requirement, as users confronted with first-party consent dialogues may be overwhelmed with information about the tens or hundreds of other third parties involved, and subjected to deceptive design patterns [5].

New governance and accountability obligations. The GDPR introduces new governance and accountability obligations on data controllers. This includes mandatory breach notifications (Articles 33 and 34), record keeping of processing activities (Article 30), data protection officers at larger companies (Articles 37–39), explicit obligations for data processors (Articles 28 and 29), and data protection impact assessments (Article 35). More generally, the GDPR puts forward the principles of data protection by default and design (Article 25), which aim to

make data protection an integral part of any personal data processing. These new obligations may be much harder for third-party trackers to meet in practice, for example where record keeping of individuals' consent is impossible due to their technical configuration [141].

Improved enforcement mechanisms. To ensure compliance, the GDPR enables large fines for violations of data protection provisions, of up to €20 million or up to 4% of total global annual turnover (whichever is higher). Further, the GDPR has a global reach: All companies operating in the EU (even those based outside the EU who are processing EU citizens' data) must comply with it (*Lex loci solutionis*). The law also seeks to reduce legal fragmentation among EU member states. As a common legal framework for data protection, the GDPR enables the exchange of personal data across the 27 EU member states, thereby allowing businesses to exchange data supposedly seamlessly. Additionally, the GDPR allows for the propagation of personal data beyond member states to countries designated by the EU Commission to have 'adequate' levels of data protection. These countries currently include the UK, Canada, Japan, New Zealand, and Switzerland.

These new enforcement mechanisms are already used in practice, to reduce the tracking of individuals. The French data protection authority CNIL fined Google multiple times over violations of the GDPR [64, 142]. It also ruled against the practices of the French advertising company Vectaury [141]. The UK data protection authority ICO investigated the legality of real-time bidding advertising, and stated that the data protection 'issues will [not] be addressed without intervention' [143]. Indeed, the Belgian data protection regulator has concluded in February 2022 that real-time bidding, as it is commonly integrated into websites and apps, is in violation of the GDPR [89]. While the regulatory enforcement of data protection law against tech companies was rare under the DPD [124, 144], this seems to have changed since the GDPR, with regulators targeting both smaller (e.g. Vectaury) and larger (e.g. Google) tracking companies. This may reflect not only the changes in provisions of the GDPR compared to the previous data protection regime, but also the increased powers and budgets of regulators since the introduction of the new law [145].

4.1.2 Challenges to the Effectiveness of the GDPR

Various aspects challenge the GDPR's effectiveness in practice. The GDPR has led to a proliferation of deceptive and arguably meaningless consent banners online [5, 7]. Such banners often violate the strict principles for consent under the GDPR and make users' consent process more complicated than intended by the GDPR's transparency principles (Article 5), but the enforcement of the law remains difficult. While growing, many regulators still operate on tight budgets [124, 145], and the *one-stop-shop* principle of the GDPR incentivises tech firms to set up their headquarters in member states with relatively lax enforcement. For instance, the Irish Council for Civil Liberties found in 2021 that Ireland is the 'bottleneck of GDPR enforcement against Big Tech' because it failed to resolve most major cases against these tech companies [146]. The Age Appropriate Design Code introduced by the UK ICO in September 2021, as a clarification of the GDPR's requirements for children (GDPR-K) in the UK, made explicit requirements for online tracking of children's data against their best interests. However, proving the (non-)existence of tracking activities and their impact on children is expected to be challenging for both technology innovators and law enforcement.

The GDPR is also *technology-neutral*, which can make it difficult for practitioners to translate the GDPR's requirements into software [127, 128]. Smaller companies that lack sufficient legal expertise or compliance budgets (e.g. independent app developers) struggle to implement the GDPR [30, 32]. Furthermore, the GDPR does not contain direct obligations for software developers [127], and the allocation of responsibility for data processing remains a topic of contentious debate. This is why Giannopoulou argued that 'more focus should be placed on the level at which privacy design decisions are truly taken and that is at an infrastructural level currently not taken into consideration within the accountability structure of the GDPR' [147]. Especially in the tracking ecosystem, a small number of tracker companies develop the dominant tracking technologies, and ship these to app developers in the form of premade tracking libraries. App developers usually

neither have access to the corresponding source code nor have a say in how these technologies are developed, and according to whose interests and values [30].

Paradoxically, the GDPR might actually contribute to the business models of large *ad tech* companies, by putting a market liberal ideology before the protection of personal data [148–150]. At the same time, there is growing evidence that there are indeed widespread infringements of the GDPR and other data protection laws in the app ecosystem [6, 36, 58, 70, 151, 152], as also highlighted by the work in the following Chapters.

4.1.3 Summary

The changes under the law, particularly the high potential fines, led many to expect that the GDPR would substantially change invasive data collection practices, including third-party tracking. Even though the key principles of the GDPR are similar to those of the DPD, there is reason to believe that the nature and extent of user tracking in mobile apps may have changed since the enforcement of the GDPR in 2018, in light of increased potential fines and regulatory enforcement, a higher bar for consent (which is necessary for most forms of tracking), and heightened expectations of the public (see Section 2.1.3). At the same time, the GDPR is not perfect. There remain various challenges to the law's effectiveness, particularly as to how the law integrates into established software development processes.

Our subsequent empirical investigation is not sufficient to establish whether the GDPR is causally responsible for any changes in third-party tracking, and to what extent. However, if the GDPR has indeed, as many had hoped, tackled excesses of personal data processing, we should expect at least some changes in the distribution, ownership, and concentration of third-party tracking in its wake. Further empirical work would be required to establish a causal relationship between the GDPR and such changes.

4.2 Methodology

Our methodology builds on static analysis to analyse tracking in the app ecosystem at scale. We proceed in four steps: app discovery and download (see Section 3.1.1), tracking detection, company resolution (see Section 3.1.2), and market concentration analysis. The first two steps replicate the work of Binns et al. [4] on analysing third-party tracking in nearly one million Android apps in 2017, which analysed apps from the UK app store. Since these authors shared their data and analysis tools publicly, this study can be reproduced. The last step replicates a study by Binns et al. [47] that computed the market concentration of tracking companies from 5,000 apps and 5,000 websites, but at a larger scale. In contrast to these previous studies, we focus on the changes in the tracking ecosystem over time and since the introduction of the GDPR. We summarise the limitations of our study in Section 4.4 ('Discussion').

4.2.1 Tracking Detection

To detect tracking in apps, we performed an automated scan of apps' `*.dex` files (corresponding to the compiled application code) to identify all URLs (strings starting with `http://` or `https://`). We then manually cross-referenced all URLs corresponding to hosts occurring in at least 0.1% of apps (in 2017 or 2020), to verify hosts corresponding to trackers. We used the same definition for a tracker as the previous study: *'a third-party tracker [is] an entity that collects data about users from first-party websites and/or apps, to link such data together to build a profile about the user.'* [4, p. 9]

Overall, we considered more hosts than in the initial study by Binns et al. [47]. These authors considered hosts occurring in at least 0.5% of apps in a set of 5,000 Android apps (compared to 0.1% of one million apps from 2017 and 2020 in our study). We additionally verified that our results held when considering the presence of *tracker libraries* in apps, another metric for tracking commonly studied in the literature. The use of tracking libraries is a common way for app developers to integrate tracking capabilities into their apps, because of the ease of

integration. However, the detection of tracking libraries might fail if developers use obfuscation techniques to hide their use of tracker libraries [153], or use non-standard ways to integrate tracking into their apps (e.g. linking to a Facebook fan page inside an app). This is why we opt for an analysis of tracker hosts in apps, which is more robust towards code obfuscation and the use of non-standard ways of tracking, and also more easily reproducible than past efforts to detect tracking libraries despite code obfuscation.

4.2.2 Market Concentration Analysis

A common measure for market concentration in economics is the *Herfindahl-Hirschman Index* (HHI). Given market shares $s_1, \ldots, s_N$ for N companies, the HHI is defined as

$$\text{HHI} = \sum_{i=1}^{N} s_i^2. \tag{4.1}$$

The HHI can attain values between 0 and 1: a higher HHI indicates a more concentrated market. A market with an HHI above 0.1 is considered potentially concentrated by EU competition regulators[154], and may motivate a market investigation. US competition regulators use higher thresholds.

The market share of a tracking company is not trivial to investigate. Traditionally, market share is measured in terms of a firm's share of revenue or unit sales of the industry total. However, in the context of free digital services, market share is typically defined in terms of the share of users for the service type (e.g. web browsers or search engines which are not revenue-generating or 'sold' to consumers)[48] Similarly, revenue or unit-sale based measures of market share do not translate into the mobile tracking ecosystem straightforwardly. Rather, market power in third-party tracking arises from a tracker company's ability to collate personal data across a variety of contexts and generate valuable insights as a result.

To reflect this situation, Binns et al. [47] proposed two measures to measure the market share of a tracker: the integration share (ISH) and the prominence-weighted integration share (PROWISH). The ISH measures the popularity of a tracker with *app developers*, and expresses this popularity relative to other trackers. The PROWISH measures the presence of a tracker in apps most popular with *app users*, also relative to other trackers. Using the ISH (PROWISH) in Equation 4.1 then gives the ISH-HHI (PROWISH-HHI).

The integration share (ISH) s_i of a tracker company t_i is computed from its prevalence, i.e. the percentage of apps that this company is present in:

$$s_i = \frac{\text{prevalence}(t_i)}{\sum_{j=1}^{N} \text{prevalence}(t_j)}. \tag{4.2}$$

The prevalence has been used widely across the app analysis literature, as a means to assess tracker adoption in apps. Computing the ISH s_i for every tracker company t_i, and using the computed s_i in Equation 4.1 yields the ISH-HHI.

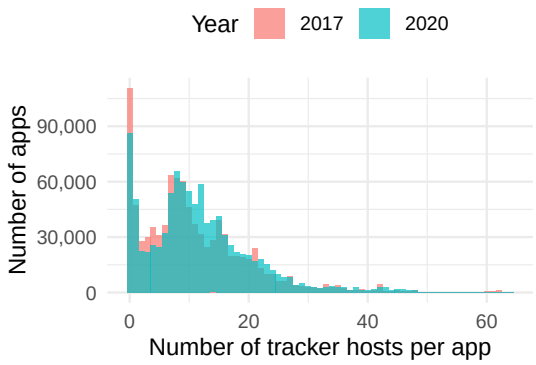
We additionally study the prominence of a tracker company t_i as the share of overall app installs that this company is present in:

$$\text{prominence}(t_i) = \frac{\text{Sum of installs of apps with tracker } t_i}{\text{Sum of all app installs}} \tag{4.3}$$

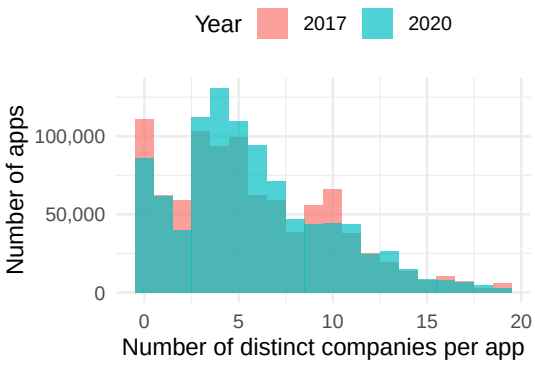
Using the prominence instead of the prevalence in Equation 4.2 gives a prominence-weighted integration share (PROWISH) s_i. Using the s_i computed from the prominence in Equation 4.1 gives the PROWISH-HHI. For a more in-depth discussion, see the original publication by Binns et al. [47].

It is important to note that we compute the PROWISH differently than in previous work [47, 155], which focused on app ranks instead of app installs. The aim of both approaches is the same: approximate the number of users that a tracker company has access to.

The assessment of market shares remains the subject of ongoing debate; so far, there has been limited intervention by competition authorities against excessive and increasing access to personal data by a single company[123].



Year	Median	Count>20	None
2017	9	14.30%	11.56%
2020	11	15.72%	8.63%



Year	Median	Count>10	None
2017	5	15.27%	11.56%
2020	5	15.76%	8.63%

Figure 4.1: Number of tracker hosts per app (left) and companies behind hosts (right) in free apps on the Google Play Store. We exclude extreme outliers having more than 65 tracker hosts (left).

4.3 Results

4.3.1 Downloaded Apps, Installs, and App Death

We downloaded a total of 1,000,750 apps between January and February 2020. This is about 2.5 years after the original study, which collected 958,270 apps between August and September 2017. Only 33.9% of the previous apps were still available on the Google Play Store in 2020; the remaining Play Store entries did not exist anymore (though they might still exist elsewhere, e.g. outside the Play Store). The median app was last updated on the Play Store in January 2017 for the 2017 data set and in June 2019 for the 2020 data set. 75.8% of 2020 apps were last updated since 25 May 2018, when the GDPR came into force.

4.3.2 Numbers of Distinct Tracker Hosts in Apps

Apps from both years contained a high number of distinct hosts in their source code that belong to tracker companies ('tracker hosts'). Their number was highly right-skewed, see Figure 4.1 (left). The median number of tracker hosts included in an app was 9 in 2017, and 11 in 2020. 14.30% of 2017 apps contained more

than 20 tracker hosts, compared to 15.72% in 2020. 88.44% contained at least one in 2017, and 91.37% in 2020, a slight increase.

4.3.2.1 Numbers of Distinct Companies behind Hosts

The prevalence of 'leaf' tracker companies (i.e. companies at the lowest subsidiary level, such as 'Flurry' as a subsidiary of 'Yahoo!' in Figure 3.1) in apps was highly right-skewed, see Figure 4.1 (right). The median number of companies was 5 in both years. 15.27% contained more than 10 companies in 2017, 15.76% in 2020.

The maximum number of companies referenced in a single app was 45 in 2017, and 43 in 2020. Among the 68 apps from both years that referenced more than 40 companies, 34 were related to photo editing, 21 to dating, 7 to sports news, 2 to games, and 1 to time tracking. This underlines how seemingly innocent apps (e.g. photo editing, time tracking) but also highly sensitive apps (e.g. dating) can expose personal data to an unexpected number of companies.

Since many tracker companies belong to a larger consortium of companies, we can also consider tracking by 'root parent' (e.g. 'Flurry' is ultimately owned by its root parent 'Verizon Communications', see Figure 3.1). Figure 4.2 shows both the 'prevalence' of root parents (i.e. the percentage of apps that contain this tracker) and their 'prominence' (i.e. the percentage of total app installs that ship this tracker).

The overwhelming share of apps included hosts belonging to Alphabet/Google and Meta/Facebook. Alphabet/Google has even increased its presence in apps slightly, while Meta/Facebook has lost some market share. Twitter has also lost some market share, and has been overtaken by Microsoft. Oracle has greatly increased its market share (especially in prominence), since its acquisition of Moat in 2017. Beyond these digital behemoths, many specialised tracking companies (including AppLovin, AdColony, Chartboost) are among the market leaders when considering their 'prominence' (i.e. share of app installs). The prominence plot also reflects the acquisition of Vungle by Blackstone in 2019, resulting in a change of root company. The median tracker company has increased its market share (prevalence and prominence both up from 3.1 in 2017 to 3.4 in 2020). Overall, the

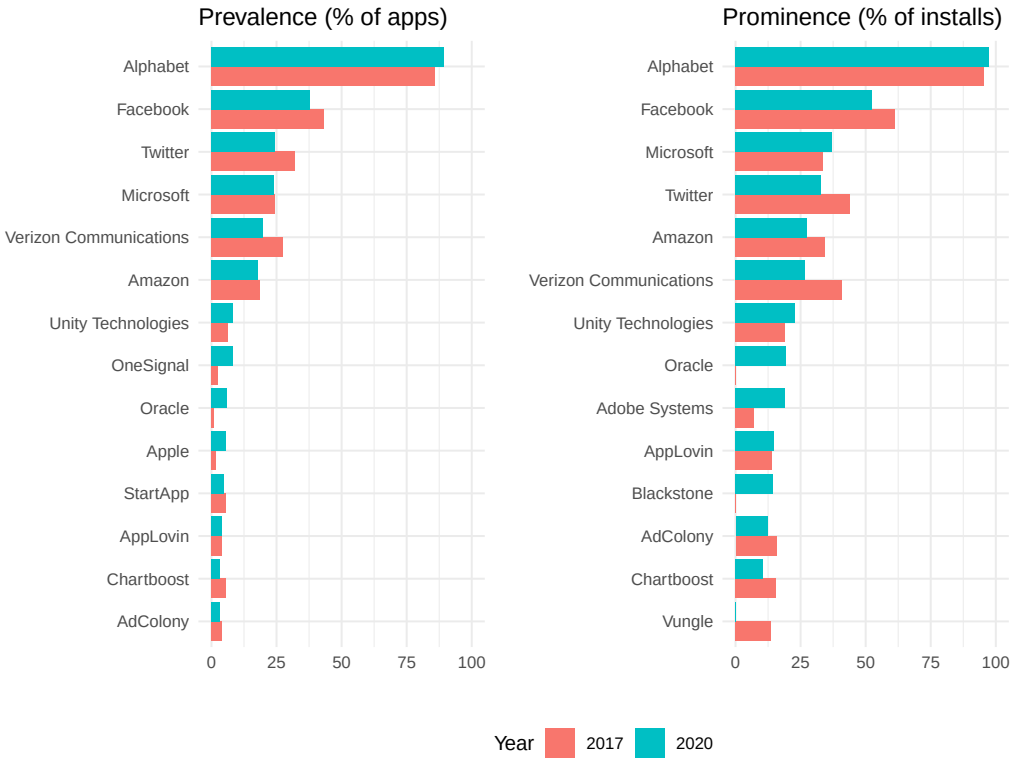


Figure 4.2: Prevalence and prominence of hosts relating to certain root tracking companies in apps in 2017 and 2020. We consider the top 11 companies from both years and for both prevalence and prominence. The companies are ranked by the values in 2020.

tracking market has seen no new entrants into the top 7 companies, both when ranking by 'prevalence' and 'prominence'.

4.3.2.2 Company Prevalence by Genre

There exists a wide range of genres on the Google Play Store to help users explore apps better. The overall number has remained at 49 since 2017. Since the genres have stayed the same, we group these genres into the same 8 'super genres' as Binns et al. [4] to provide high-level statistics about the apps (for example, the genres 'Comics', 'Sports', 'Video Players', and all games are all grouped into 'Games & Entertainment'). Children' apps are singled out (which are assigned to 'Family' categories on the Play Store), given the concern around data collection from this group of app users. We re-ran the company analysis for each super genre, see Figure 4.3.

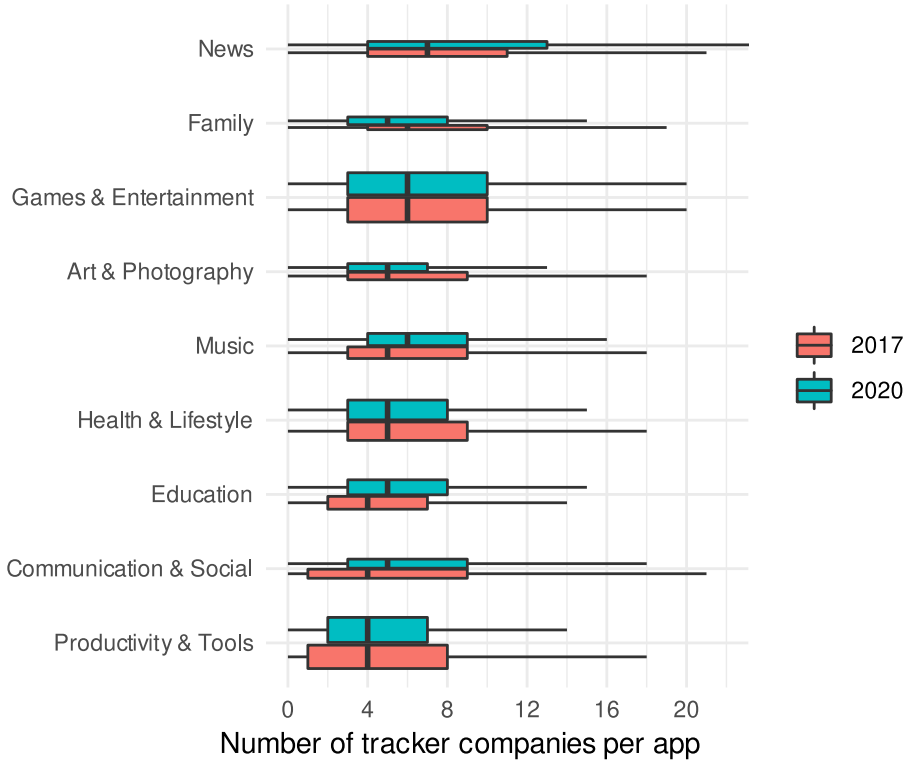


Figure 4.3: Boxplot of number of distinct tracker companies behind hosts referenced in apps, grouped by super genre. Black bars indicate medians. Height of bars indicates number of apps in a given super genre.

The genre with the most tracker companies was 'News' (seven companies), both in 2017 and 2020. Family (Children apps), having the second most tracker companies in 2017 (6 companies), was down by one company in the median. 'Music', 'Education', 'Communication & Social' were up by one company on average. Overall, the presence of tracking in apps was similar between the years across super genres.

4.3.2.3 Country Differences

We also analysed the countries tracker companies in apps are based in (including the subsidiary and all parent companies, see Table 4.1). About 90% of apps contained a tracker that is owned by a US-based company. The next most common countries were China and Russia in both years. The top 8 countries were also the same, but with different rankings. South Korea has seen an increase by 22%, while Germany and Israel have both seen a decline by about 40%. However, overall, the fluctuation between the years is small across the top 8 countries. Overall,

Country	% Apps (2017)	% Apps (2020)
US	88.38	91.31
China	6.35	6.11
Russia	4.12	4.19
Germany	4.04	2.42
South Korea	3.18	3.88
UK	2.92	2.68
India	2.11	1.83
Netherlands	1.87	1.57

Table 4.1: Apps including at least one tracker associated with a company within a given country.

the share of tracker companies based in UK and EU member states has both somewhat decreased between 2017 and 2020.

We also computed the country prevalence for each super genre. The US stayed the most prevalent country (between 83% and 95% for 2017, and 88% and 97% for 2020). China was present in about 9.2% of apps from the 'Health & Lifestyle' genre in both years.

4.3.3 Changes in Company Structure

We now analyse the network of companies involved in tracking, and how this has changed from 2017 to 2020. We included those tracker hosts occurring in more than 0.1% of apps, the tracking companies owning these hosts, and all their parent companies. We refer to all included companies as having a *significant market share* in app tracking. 0.1% might seem small, but can still amount to millions of individuals, since there are billions of Android users.

In total, there were 164 companies (including all parent companies) with a significant market share in 2017, compared to 162 in 2020. There were a total of 102 root companies with a significant market share in 2017, compared to 89 in 2020. On average, a company consortium consisted of 1.48 companies in 2017, compared to 1.65 in 2020. 62 companies were owned by another in 2017, compared to 73 in 2017. All these figures point to a subtle consolidation of tracking companies since 2017.

One straightforward explanation for the consolidation of the tracking ecosystem would be a substantial number of companies *losing a previously significant market share* (i.e. losing access to at least 0.1% of apps). Companies that have lost their significant market share include Myspace, Loggly, and BugSense. However, no larger tracking companies have been affected by this.

Another important reason for consolidation in the tracking market were *mergers and acquisitions (M&A)*. We found a total of 53 M&A transactions between the beginning of 2018 and June 2020 among tracker companies. For instance, Blackstone, one of the largest investment firms, entered the tracking market with the purchase of the advertising firm Vungle in July 2019. Media Games Invest, another investment firm, purchased the tracking companies PubNative and Verve, as part of over 30 strategic acquisitions over the past six years (Gardt, 2020). Verve, in turn, had purchased the advertising company Receptiv in May 2018. Overall, there were 7 investment firms in our company data set with a significant market share.

Three of the 53 observed M&A transactions were filed with EU or UK competition authorities: Bain Capital Investors / Kantar, Silver Lake / ZPG, and Taboola / Outbrain. The first two were filed with the European Commission, which did not pursue in-depth investigations and approved the M&A transactions within a few weeks. Taboola filed the planned acquisition of its rival Outbrain with the UK Competition and Markets Authority (CMA) in April 2020. The CMA opened a phase 1 investigation, and found potential competition concerns leading to a phase 2 investigation from June 2020. Taboola eventually abandoned its acquisition plans in September 2020, which made the CMA cancel its investigations.

We have also observed 11 rebrandings among prevalent tracking companies. For example, Verizon Communications has restructured its media operations internally, inside its subsidiary Verizon Media (previously known as Oath). Amazon renamed its 'Amazon Marketing Services' to 'Amazon Advertising', thereby seemingly trying to advance its mobile advertising business. Amazon also purchased the advertising firm Sizmek in 2019, after a three-year ownership by the private equity firm Vector Capital led to bankruptcy. Microsoft has integrated BitStadium (purchased in 2014)

Year	ISH-HHI	PROWISH-HHI	Gini
2017	0.112	0.071	0.491
2020	0.115	0.067	0.493

Table 4.2: Market concentration and equality measures. All three metrics take values between 0 and 1. A higher HHI value indicates more concentration in the tracking market. A lower Gini coefficient indicates higher equality between the market participants.

into its other cloud services, and rebranded it as 'Microsoft App Center'. Notably, after our data collection, Facebook rebranded itself as 'Meta'.

4.3.4 Market Concentration

We now consider how the market concentration of tracking companies has changed between 2017 and 2020. As discussed in our methodology in Section 4.2, we use two metrics: the ISH-HHI and PROWISH-HHI. The results can be seen in Table 4.2.

The ISH-HHI has seen a subtle increase, and the PROWISH-HHI a subtle decrease. Since the ISH-HHI is in the range of 0.1, this shows some signs of concentration in the integration of tracking into all apps in both years. However, when weighing apps by their prominence (i.e. by the number of app installs), the concentration decreases to about 0.07. The Gini coefficient, an inequality metric herein computed among root tracking companies, has increased subtly (see Table 4.2). This suggests a slightly decreased equality in terms of market access of tracker companies in 2020.

Overall, there has been very limited change across all studied market concentration measures.

4.4 Discussion

In this Section, we discuss the above findings in the context of our original research questions, and their implications for the ongoing development of data rights regulation and the future of the third-party tracking sector.

The distribution of third-party trackers has not changed much. Our results suggest that the GDPR has not had a large effect on the distribution of third-party tracking across apps on the UK Google Play Store. The same handful of third-party tracking companies have similar prevalence and prominence; the average app contains a similar number of third-party trackers (measured at the level of companies rather than hosts); and a consistent percentage of apps (15%) contain more than ten trackers. If the GDPR has led to changes in tracking practices, they are not showing in the distribution of trackers. This might seem surprising, given that the GDPR and ePrivacy present challenges for compliance in the context of multiple third parties. Rather than reduce the number of third parties they share data with, to enable compliance with the requirements of consent, record-keeping, data protection by design, transparency and accountability, first-party app developers continue to share data with multiple third parties.

Some small changes in the distribution of third-party trackers have been observed. Alphabet-owned trackers have slightly increased in both prevalence and prominence, while others such as Meta/Facebook and Twitter have decreased on both measures. The number of apps with no trackers at all has decreased from 11.6% to 8.6%. While these might in some way be indirect effects of the GDPR, we find no clear explanation connecting them.

Cross-jurisdictional data flows. As explained above, the EU data protection regime enables the free-flow of personal data across EU member states, the UK and other countries that are deemed to provide 'adequate' data protection standards, as determined by the European Commission. This does not mean that data being sent to a third-party tracker based outside the EU / UK's list of adequate countries is necessarily unlawful; some tracker companies may designate local subsidiaries as the data controller for the personal data of citizens in the EU / UK, and transfers to non-adequate countries may still be lawful with the use of alternative measures including 'standard contractual clauses' (Article 46 GDPR) and 'binding corporate rules' (Article 47 GDPR).

Given that these alternative options come at substantial costs, it would be reasonable to expect at least some decrease in the number of third-party trackers based in non-adequate countries, as first parties seek to minimise the compliance risk of unlawfully transferring data across borders. However, despite their jurisdiction not being deemed as 'adequate', companies based in the US, India, China, and Russia were still behind a large portion of the tracking observed in our analysis in 2020. In particular, organisations based in the US (about 90%) and China (about 9%) led the pack for third-party tracking in the 'Health & Lifestyle' super genre. These findings are potentially concerning: absent specific justifications, the GDPR prohibits processing data concerning health (see Article 9(1) GDPR). While our study did not determine which, if any, third-party trackers were collecting data that could be treated as health-related data under the GDPR, there is often a risk of accidental disclosure of sensitive information (e.g. the information that an individual uses certain sobriety or mental health apps) (Norwegian Consumer Council, 2020). Overall, there has been limited change in sending data to trackers in non-adequate countries (which includes the US). Indeed, there is a slight reduction in the prevalence of third-party trackers based in significant countries *inside* the EU (Germany and the Netherlands), supporting the claim that GDPR may actually be helping global tech firms *outside* the EU[149].

Market concentration and competition. Our analysis hints at a high level of concentration in the tracking market. Alphabet/Google and Meta/Facebook continue to dominate app tracking. Their dominance is particularly present in the number of apps they cover. If these companies can show ads on devices that other competitor advertising companies hardly have access to (e.g. due to the *default bias* of app developers to use the software solutions of established brands [30, 31]), they can extract sizeable revenues from their dominance of the tracking market, and might even be able to exert meaningful control over advertising prices [48]. From our data, this seems to be particularly the case for those apps that have few installs, but represent the vast majority of apps on the Play Store due to the long-tailed distribution [4, 76].

At the same time, many relatively smaller companies are involved in app tracking. Some of these manage to reach fairly high market shares in terms of app installs (including AppLovin, AdColony, and Chartboost). These smaller companies usually focus exclusively on mobile advertising, instead of having a broad portfolio of digital services like Alphabet/Google, Meta/Facebook, or Verizon. The specialisation and small size of these tracking companies seems to allow them to gain a certain competitive advantage, and potentially offer better deals to app publishers (who might otherwise choose the market leaders). An important competitive advantage of these companies might be reduced public awareness and regulatory scrutiny, allowing them to compete with the market leaders in certain segments, at the expense of data protection and user privacy.

Smaller companies may have access to fewer apps, but they might still be able to gain deep insights into the lives of individuals, especially at the aggregate level. Even if a tracker company gets access to a small subset of users only, the use of *permanent user identifiers* can enable these companies to exchange data with other tracking companies, such as data brokers, and gain insights into larger numbers of users. The average user has about 30 apps installed at any given time [156, 157], but for a third-party aiming to obtain a profile of the user, it might be sufficient to be integrated into only one of those apps. As such, there may be diminishing returns for third parties aiming to increase their prevalence or prominence in the app marketplace.

While a concentration of data with only a few companies can help transparency of tracking and compliance with data protection and privacy legislation, it also puts more power into the hands of a few companies. By contrast, a tracking ecosystem with dozens of market participants – as we continue to have – is difficult to oversee by regulators and the interested public.

4.5 Limitations

The analysis in this Chapter has certain limitations. The analysis of hosts in apps only gives a partial picture of app tracking, as explained in Section 4.2.1. We do

not analyse the handling of personal data on the servers of tracking companies or how these companies might share data with other companies, but only tracking that happens directly on users' devices. We only focus on tracker hosts that are present in at least 0.1% of apps. Some of these hosts may never be contacted, while other hosts may not be present in the app code at install time. Further, the definition of 'tracking' (see Section 4.2.1) is, while based on the protocols of previous research by Binns et al. [47], open to debate. Lastly, we treat all tracker hosts equally, and do not account for different purposes (e.g. advertising and analytics) or different levels of intrusiveness. While this Chapter focuses on Android apps and the Google Play Store, tracking is also widespread on iOS and the Apple App Store as we will see in Chapter 6.

4.6 Conclusions & Future Work

In this Chapter, we analysed the presence of third-party tracking in apps, before and after the introduction of the GDPR. We found that tracking has remained prevalent across a wide range of mobile apps and prominent in its reach of app user data. The number of tracking companies has stayed about the same between 2017 and 2020 in the average app on Google Play. The top destination countries have likewise stayed the same, as have the most prominent tracking companies – namely Alphabet/Google and Meta/Facebook – and the sending of personal data to trackers based in a third-party country without an 'adequate' level of data protection. Our observations are consistent across *super genres*. Apps continue to rely on tracking technologies, e.g. to retrieve analytics and show advertising, even after the introduction of the GDPR. The law does not appear to have changed these incentive structures fundamentally.

We also found that the market concentration in the tracking ecosystem has seen limited change over time. Competition between tracking companies seems to revolve at least partly around data protection and user privacy due to the relevance of little-known tracking companies that evade public and regulatory scrutiny but

collect data about sizeable numbers of individuals. As such, our study provides empirical evidence of fears expressed in previous academic work that the GDPR might entrench the existing power imbalances in the digital ecosystem [148–150].

While our current analysis points to limited change in the tracking ecosystem so far, change might be imminent. Apple and Google have been introducing various privacy measures that could, despite increasing the concentration of data collection with these companies, improve data protection and user privacy. The most notable recent example is Apple's introduction of mandatory user opt-ins to tracking in iOS apps in April 2021. First reports suggest high refusal rates of tracking [158, 159], with the direct result of tripling the iOS market share of Apple's own advertising business [160], which itself sidesteps the new rules against tracking [161]. However, the effects of this new policy are still subject to ongoing debate and analysis. Meanwhile, Google is considering removing third-party cookies from its Google Chrome browser and replacing them with the Topics API (formerly FLoC), thereby shifting away from identifying individuals to targeting cohorts of users with similar interests. Google is also considering similar steps within apps with its *Android Privacy Sandbox*. We will explore the impact of Apple's changes more in Chapter 7.

An important driver of these new privacy measures has been the emergence and overhaul of data protection and privacy laws around the globe, more extensive regulatory action, and ultimately the increased privacy expectations of citizens. In this sense, the GDPR has already contributed to changing the mobile tracking ecosystem by shaping people's expectations around privacy and increasing data protection enforcement. Beyond the EU, the GDPR has also encouraged the emergence of new and revised data protection laws, notably in Brazil, Japan, China and California. In the UK, the government is discussing a reform of its domestic implementation of the GDPR. Meanwhile, the EU is planning to introduce a new ePrivacy Regulation, which would overhaul and supersede the existing ePrivacy Directive in the EU, but not in the UK, leading to further regulatory divergence. According to our analysis, the lack of enforcement of the existing rules is one of the key issues that needs to be addressed.

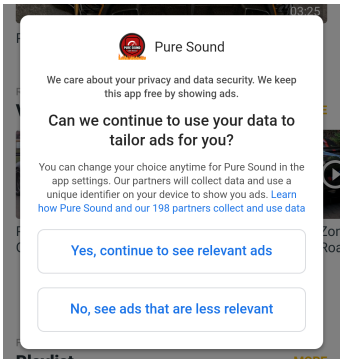
Transparency is essential in holding power to account, but the analysis of privacy practices remains difficult in the mobile tracking ecosystem. This conflicts with the strict transparency requirements for the processing of personal data laid out in the GDPR (Article 5). More research as well as changes to the current data protection and privacy practices of the gatekeepers will be needed to afford regulators and independent researchers more transparent access and to build more sustainable business models that can live without the continuous surveillance of those individuals that these technologies are meant to serve. The following Chapters will try to provide more detailed insights into the data practices around tracking, and also develop the necessary methodology to support these efforts with ease.

5
Consent to App Tracking

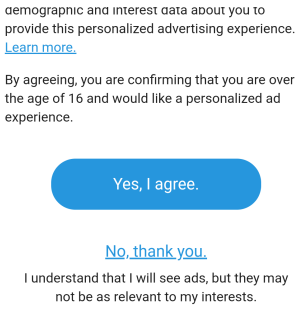
Data protection and privacy legislation such as the General Data Protection Regulation (GDPR) [162] in the EU and the UK, and the Children's Online Privacy Protection Act (COPPA) [163] in the US, establish clear rules when it comes to the processing of personal data and provide additional safeguards when it comes to information relating to children. As explained in Section 5.2, consent is usually a necessary precondition for third-party tracking. The implementation of consent has even, as explained in Section 2.3.2, sparked a fierce public battle between Apple and Facebook over tracking controls in iOS 14.5 around the end of 2020 [105, 106]. The importance of consent aside, there exists little empirical evidence of whether mobile apps implement any type of consent mechanisms before engaging in tracking.

Driven by these observations, this Chapter aims to answer the following research questions:

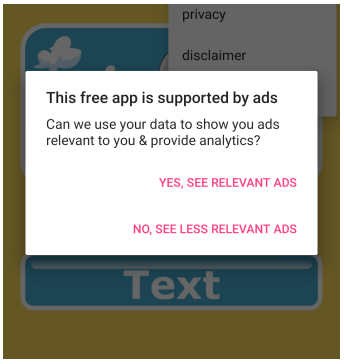
1. Do app developers need to obtain valid user consent before engaging in third-party tracking in the EU and UK? *(consent requirements for tracking)*
2. To what extent do apps engage in third-party tracking, and obtain valid user consent before doing so? *(practices of app developers)*



(a) This app uses the Consent API developed by Google. The popup suggests that personal data may be shared with 199 companies before user consent is given ('continue').



(b) This app uses the consent implementation by Twitter MoPub. By declining, a user rejects a 'personalized ad experience', but potentially not all app tracking.



(c) This app uses a custom consent solution. Consent is not granular. The answer options do not match the question. It is unclear if 'No' rejects analytics.

Figure 5.1: While most apps on the Google Play Store use third-party tracking, only few apps allow users to refuse consent (less than 3.5%). The figure shows three common examples of these 3.5% of apps. Since very few apps give users a genuine choice over tracking, this Chapter suggests widespread violations of EU and UK privacy law.

3. To what extent do third-party tracking companies encourage and support app developers to obtain consent as and where required? *(practices of tracker companies)*

We previously explored in Section 2.1 that individuals often struggle to express their choice concerning their data use – termed the *privacy paradox*. Despite such limitations, consent remains a key component of many privacy and data protection regimes. For the purpose of this Chapter, we do not assume that consent is the only or best way to address privacy and data protection issues. Rather, we aim to investigate whether, in addition to all these problems and limitations, the basic process of consent itself is even being followed where it is currently required in the context of third-party tracking in apps. In this Chapter, we also use a rather broad definition of consent, classifying any *affirmative action* as consent; we explain more about this methodological choice in Section 5.3.

Contributions. In answering these questions, this Chapter makes three contributions. First, we clarify the role of consent in the regulatory framework

applicable in the EU and the UK when it comes to the processing of personal data for third-party tracking. Second, we provide empirical evidence as to a widespread absence of consent mechanisms to legitimise third-party tracking in 1,297 apps. Third, we analyse the guidance provided by 13 commonly used tracker companies and assess whether they inform app developers about how to translate consent in code (see Figure 5.2 and Table 5.2).

Structure. The rest of this Chapter is structured as follows. Section 5.1 reviews the existing system-wide tracking controls for Android. Section 5.2 discusses the role of consent for third-party tracking in the EU and UK by drawing on the guidance issued by national Data Protection Authorities (DPAs). Section 5.3 analyses the presence of consent for third-party tracking in 1,297 Android apps randomly sampled from the Google Play Store. Section 5.4 reviews the guidance offered by tracker companies to app developers. We first discuss our results in Section 5.5, then turn to the limitations of our approach in Section 5.6 and our conclusions in Section 5.7.

5.1 Alternatives to In-App Consent

Before turning to the legal analysis concerning when consent for third-party tracking within individual apps is required, it is worth considering the options users currently have to limit app tracking on Android at a system level. This is pertinent to our subsequent analysis because, if system-level controls were sufficient, the question of efficacy and compliance with individual app-level consent requirements might be redundant. The options for users fall into three categories: system settings, system modification, and system APIs.

System settings. The Android operating system offers users certain possibilities to limit unwanted data collection. Users can manage the types of data each app can access through *permissions*. This does not stop tracking, but blocks access to certain types of data, such as location. A problem inherent in the permission approach is that trackers share permission access with the apps they come bundled with. This means that, if a user allows location access to a maps app with integrated

trackers, all these trackers have access as well. This, in turn, might give users a false sense of security and control. Google offers users the possibility to opt-out of personalised advertising. If users choose to do so, apps are encouraged to cease using the system-wide *Google Advertising Identifier* (AdID) for personalised advertising (although apps can continue to access the AdID). Unlike iOS, Android does not offer the option to opt-out of analytics tracking using the AdID, or to prevent apps from accessing this unique user identifier on all such devices. However, Google intends to change this [164].

System modification. Since the early days of Android, many developers have set out to modify its functionality and implement better privacy protections. *Custom ROMs* are modified versions of Android that replace the default operating system that comes pre-installed on Android smartphones. Popular examples are Lineage OS and GrapheneOS, which both try to reduce the dependency on Google on Android and increase user privacy. Another is TaintDroid, which monitors the flow of sensitive information through the system [66]. A popular alternative to custom ROMs is *rooting* devices by using exploits in the Android system to gain elevated access to the operating system, or by changing the bootloader of the Android system. Rooting is a necessary prerequisite for many privacy-focused apps, including AdAway [165], XPrivacy [166], and AppWarden [167]. System modification grants maximum control and flexibility regarding tracking, but requires a high level of technical expertise. It also relies on security vulnerabilities, often creating risks for (non-expert) users.

Google nowadays restricts attempts to modify Android by preventing custom ROMs from running apps using *Google's Safety Net*. This is meant to protect sensitive apps (e.g. banking apps) from running on unsafe devices, but is also used by other popular apps such as Pokemon GO and Snapchat [168]. Some Internet outlets have declared the 'end for Android rooting, [and] custom ROMs' [169].

System APIs. Another alternative to system modification is to develop apps that build on the capabilities of Android's system APIs to detect and block network traffic related to tracking. This is possible without the need for system modification at the cost of more advanced functionality. Popular apps in this category include

AdGuard (using a local VPN on the Android device) [170] and DNS66 (changing the DNS settings of the Android device) [171]. Another is NetGuard [136], a firewall that allows users to monitor network connections through a VPN, and to block certain domains manually. All these tools block connections regardless of the actual content of the communications. These content-agnostic approaches can lead to over-blocking and breakage within apps.

Alternative tools aim for more fine-grained protection by removing sensitive information from network requests, such as device identifiers or location data [67, 172]. Unfortunately, these content-based approaches rely on breaking secured network connections, and on installing a self-signed root certificate on the user's device. This practice was banned by Google with the introduction of Android 7 in 2016 because of the security risks it entails [173]. While these apps grant users the possibility to block tracking through system APIs, Google does not allow them on the Play Store [103]. Instead, users must sideload them onto their devices from alternative sources, such as GitHub and F-Droid.

In conclusion, while there exists a wide array of options for end-users to reduce tracking, none of them can provide the granularity of consent implemented inside each app. Many of the existing tools require a high level of technical expertise, including root access or modifications to the operating system, and are therefore unsuitable for non-expert users. This makes many users dependent on the privacy solutions offered by apps themselves and their operating systems.

5.2 When is Consent to Tracking Required?

In this Section, we analyse whether consent is a prerequisite for third-party tracking under EU and UK law, as well as its role under the Google Play Store policy. As discussed in Section 2.4, two main legal instruments are relevant to the issue of consent to third-party tracking on mobile apps: the GDPR and the ePrivacy Directive[1].

[1]It is worth noting that the ePrivacy Directive is currently under revision. A change to the current regulatory requirements not expected soon in practice due to the nature of the EU

5.2.1 GDPR and the Need for a Lawful Ground

As covered in Section 2.4, the GDPR imposes a wide array of obligations on so-called *controllers*, paired with high fines for non-compliance, besides granting individuals various rights regarding the protection of their personal data. For the purpose of this Chapter, we assume that app developers qualify as *controllers*. In other words, this means that they 'determine the purposes and the means of the processing of personal data' (Article 4(7) GDPR). While this might well be the case when the company actually processing the personal data at stake is also in charge of the development of the app, it is important to highlight that controllership does not always end up on their shoulders. This is the case, for instance, when a company outsources the development of its app to an external team of software developers working based on clear-cut specifications and requirements, in which case the latter is likely to be considered as a *processor* (Article 4(8) GDPR) or a *third party* (Article 4(10) GDPR).

If app developers want to collect personal data for whatever purpose, they need to rely on one of the six *lawful grounds* listed in Article 6(1) GDPR. Only two usually apply in the context of mobile apps, namely: *consent* and *legitimate interests*[2]. On the one hand, and as specified in Article 4(11) GDPR, valid *consent* is

> *any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her*

As clarified by the Court of Justice of the European Union in 2020, this bans the use of pre-ticked boxes to gather consent [174]. *Legitimate interest*, on the other hand, is a viable alternative to consent but requires a careful balancing exercise between the controller's interests in processing the personal data and the data subjects' interests and fundamental rights (Article 6(1)f GDPR) [175]. The other guarantees stemming from the GDPR (including transparency, security, purpose

legislative process.

[2]The remaining four lawful grounds listed in Article 6(1) GDPR are the fulfilment of a *contract*, a *legal obligation*, the data subject's *vital interests*, and the performance of a *public task*.

and storage limitation, and data minimisation) remain applicable regardless of the lawful ground used to legitimise the processing.

Legitimate interest and direct marketing. The GDPR explicitly foresees in Recital 47 that the 'processing of personal data for direct marketing purposes may be regarded as carried out for a legitimate interest.' In the context of mobile apps, *direct marketing purposes* usually refer to the provision of mobile advertising in the app. In other words, the text of the GDPR underlines that controllers can have a legitimate interest in showing mobile ads – a success that had been celebrated by the advertising lobby at the time. Since app tracking underpins many forms of mobile ads, this Recital might support the legitimate interest of controllers in such data collection.

Consent for high-risk data processing. Despite these existing provisions in the GDPR, it is less clear what data practices are expressly permitted. Do controllers only have a legitimate interest in contextual advertising (which relies on personal data to a lesser extent), or are invasive data practices like real-time bidding advertising also covered? While the controller's legitimate interests could potentially be a viable option for legitimising third-party tracking on mobile apps, this processing is also likely to qualify as a *high-risk data processing activity*.[3] Features of third-party tracking, that indicate such high-risk processing, include the use of 'evaluation or scoring', 'systematic monitoring', 'data processed on a large scale', 'data concerning vulnerable data subjects', or 'innovative use or applying new technological or organisational solutions'. Some of these features undoubtedly apply to third-party tracking, since tracking companies usually engage in large-scale data collection, at a high frequency, across different services and devices, with limited user awareness.

[3]The Article 29 Working Party – an EU body to provide guidance on data protection law (now the European Data Protection Board) – has listed the 9 features commonly found in such high-risk activities, namely: 1) Evaluation or scoring, 2) Automated-decision making with legal or similar significant effect, 3) Systematic monitoring, 4) Sensitive data or data of a highly personal nature, 5) Data processed on a large scale, 6) Matching or combining datasets, 7) Data concerning vulnerable data subjects, 8) Innovative use or applying new technological or organisational solutions, and 9) Prevention of data subjects from exercising a right or using a service or a contract. [176]

The Information Commissioner's Office (ICO) – the UK's DPA – discourages the use of legitimate interest for high-risk activities and recommends that controllers instead rely on another lawful ground such as consent [177]. Similarly, after having analysed the case of tracking deployed on a webshop selling medical and cosmetic products [84], the German DPA concluded that the average website visitor could not reasonably expect tracking of their online activities to take place, especially when it operates across devices and services. In that case, it argued, the website visitor is not in a position to avoid the data collection. These are the concrete manifestations of the balancing exercise required by Article 6(1)f. 'Share First, Ask Later (or Never?) Studying Violations of GDPR's Explicit Consent in Android Apps' [178] further argued that data sharing from one controller to another in app tracking (i.e. from one app developer to a tracking company) would always require consent, referencing the opinion of the Article 29 Working Group on Legitimate Interest from 2014 [175]. All in all, the aforementioned considerations might disqualify the use of the controllers' legitimate interests as an appropriate lawful ground to legitimise third-party tracking in mobile apps.

5.2.2 ePrivacy and the Need for Consent for Local Storage of and Access to Data

Whether and how app tracking might be permissible under the GDPR is an important legal discussion that has not been settled yet. This represents an ongoing debate within academia and jurisprudence, and it will take more time to be finally resolved. However, there exists another law that puts forward much clearer and stricter rules around tracking: the ePrivacy Directive. This is a *lex specialis*, meaning that, when both the ePrivacy Directive and the GDPR apply in a given situation, the rules of the former will override the latter. This is the case for third-party tracking, since Article 5(3) of the ePrivacy Directive specifically requires consent for *accessing or storing* non-technically necessary data on a user's device. It is widely accepted, and reflected in DPAs' guidance, that most tracking activities are not technically necessary, and therefore require consent to read or store data on a user's device [179].

5. Consent to App Tracking

If tracker software involves accessing or saving information on a user's smartphone – as third-party trackers typically do on a regular basis – this requires prior consent. Tracking software usually *accesses* various identifiers (e.g. the Android ID, Android Advertising Identifier, the IMEI, the UDID, the phone number and name) from a user's smartphone and transmits them to tracking companies. Furthermore, as we observed in our experiments, most tracking software additionally generates a random, unique user identifier on the first app start and *stores* this identifier in the internal storage of the device. Even if tracking software operated without any such identifiers and used device fingerprinting instead, it would still need to access information (e.g. installed apps, local IP address, free device storage, battery level, volume level, MAC addresses of nearby devices) from the user's device for such fingerprinting. Since tracking – by definition – relies on the singling out of individuals, it cannot function without *some* form of identification. As a result, while consent is already the most reasonable option under the GDPR, it becomes the only viable one when combining both regulatory frameworks.

Recent guidance and enforcement action from various DPAs have also demonstrated how the GDPR and the ePrivacy requirements apply to situations where consent is the basis for processing by one controller, and when that data is provided to another controller for further processing. Article 7(1) of the GDPR requires that, where consent is the lawful ground, the controller must be able to *demonstrate* that the data subject has consented. The ICO's guidance states that third-party services should not only include contractual obligations with first parties to ensure valid consent is obtained, but 'may need to take further steps, such as ensuring that the consents were validly obtained' [180]. It notes that, while the process of getting consent for third-party services 'is more complex', 'everyone has a part to play' [180]. The responsibility of third parties has been further illustrated in an enforcement action by the CNIL (the French DPA), against Vectaury, a third-party tracking company [141]. This showed how the validity of consent obtained by an app developer is not 'transitive', i.e. does not carry over to the third party. If a first party

obtains consent 'on behalf' of a third party, according to a contract between the two, the third party is *still* under the obligation to verify that the consent is valid.

To summarise the implications of GDPR and ePrivacy in the context of third-party tracking: consent is typically required for access to and storage of data on the end-user's device, which in turn is usually required for tracking. Even if that consent is facilitated by the first party, third parties must also be able to demonstrate the validity of the consent for their processing to be lawful on that basis.

5.2.3 Requirements of the Google Play Store

In addition to EU and UK privacy law, Google imposes a layer of contractual obligations that apps must comply with. These policies apply worldwide – so beyond the jurisdiction of the EU and UK – and might oblige all app developers to implement adequate mechanisms to gather consent for third-party tracking. Google's *Developer Content Policy* highlights that in-app disclosure and consent might need to be implemented when 'data collection occurs in the background of your app' [181]. The Developer Content Policy also requires that developers abide by all applicable laws. It is unclear how strictly compliance with these policies – and in particular with all applicable laws – is verified and enforced by Google.

5.3 Tracking in Apps before and after Consent

The previous Section established that third-party tracking in apps typically requires valid user consent under the EU and UK regulatory framework – because of the ePrivacy Directive in conjunction with the GDPR. Despite these legal obligations, it is yet not clear how and whether consent is realised in practice. In order to examine the extent to which regulation around consent is implemented in practice, we conducted two studies – Study 1 (in this Section) to see how consent is implemented in a representative sample of Google Play apps, and Study 2 (in the following Section 5.4) to examine how app developers were supported and encouraged to implement consent by the providers of tracker libraries.

5.3.1 Methodology

We studied a representative sample of 1,297 free Android apps. This sample was chosen randomly (through random sampling without replacement) from the apps already analysed in the previous Chapter 4. The selected apps were run on a Google Pixel 4 with Android 10. Each app was installed, run for 15 seconds, and then uninstalled. We did not interact with the app during this time, to record what companies the app contacts before the user can be informed about data collection, let alone give consent. During app execution, we recorded the network traffic of all tested apps with the TrackerControl app (see Section 3.2.2). We did not include any background network traffic by other apps, such as the Google Play Services. For apps that showed full-screen popup ads, we closed such popups, and took note of the presence of display advertising. We assessed whether each contacted domain could be used for tracking and, if so, to what tracking company it belonged, using a combination of the App X-Ray [4] and `Disconnect.me` [182] tracker databases. 15 seconds after having installed the app, we took a screenshot for further analysis, and uninstalled it.

We inspected the screenshots for any form of display advertising, privacy notice or consent. We took note of any display advertising (such as banner and popup ads) observed. We classified any form of information about data practices as a privacy notice, and any *affirmative* user agreement to data practices as consent. While this definition of consent is arguably less strict than what is required under EU and UK law, this was a deliberate choice to increase the objectivity of our classification, and provide an upper bound on compliance with EU and UK consent requirements. We then re-installed and ran those apps that asked for consent, granted consent, and repeated the network capture and analysis steps above, i.e. monitoring network connections for 15 seconds, followed by a screenshot, and finally, removed the app once again.

Hosts	Company	Apps
adservice.google.com	Alphabet	19.7%
tpc.googlesyndication.com	Alphabet	17.2%
lh3.googleusercontent.com	Alphabet	14.2%
android.googleapis.com	Alphabet	12.9%
csi.gstatic.com	Alphabet	11.6%
googleads.g.doubleclick.net	Alphabet	10.3%
ade.googlesyndication.com	Alphabet	9.7%
connectivitycheck.gstatic.com	Alphabet	9.5%
config.uca.cloud.unity3d.com	Unity	7.5%
ajax.googleapis.com	Alphabet	6.9%
api.uca.cloud.unity3d.com	Unity	6.8%
android.clients.google.com	Alphabet	6.7%
gstatic.com	Alphabet	5.8%
graph.facebook.com	Facebook	5.5%

Table 5.1: Top contacted tracker domains by 1,201 randomly sampled apps from the Google Play Store, at launch, before any interaction with the apps.

5.3.2 Results

Of the 1,297 apps, 96 did not show a working user interface. Some apps did not start or showed to be discontinued. Other apps did not provide a user interface at all, such as widgets and Android themes. We therefore only considered the remaining 1,201 apps. 909 apps (76%) were last updated after the GDPR became applicable on 25 May 2018.[4] On average, the considered apps were released in August 2018 and last updated in December 2018. All apps were tested in August 2020, within a single 24-hour time frame.

Widespread tracker use. Apps contacted an average of 4.7 hosts each at launch, prior to any user interaction. A majority of such apps (856, 71.3%) contacted known tracker hosts. On average, apps contacted 2.9 tracker hosts each, with a standard deviation of 3.5. The top 10% of apps contacted at least 7 distinct hosts each, while the bottom 10% contacted none. Alphabet, the parent company of Google, was the most commonly contacted company (from 58.6% of apps), followed by Facebook (8.2%), Unity (8.2%), One Signal (5.6%), and Verizon (2.9%). Apps

[4]It is worth noting, however, that both the need for a lawful ground – an obligation under Directive 95/46 – and the consent requirement for access to and storing on terminal equipment – an obligation under the ePrivacy Directive – were already applicable before 25 May 2018. The latter has merely provided clarification on the conditions for consent to be valid.

that we observed showing display ads contacted a significantly higher number of tracker hosts (on average 6.0 with ads vs 2.2 without).

The dominance of Google services. The 9 most commonly contacted domains all belong to Google; the top 2 domains are part of Google's advertising business (`adservice.google.com`, linked to Google's Consent API, and `tpc.googlesyndication.com`, belonging to Google's real-time advertising bidding service). 704 apps (58.6%) contacted at least one Google domain; the top (Google) domain was contacted by 236 apps (19.7%). Such breadth and variation is reflective of the corresponding variety of services that Google offers for Android developers, including an ad network (Google AdMob), an ad exchange (Google Ad Manager, formerly known as DoubleClick), and various other services. Domains by other tracker companies, such as Unity and Facebook, were contacted less frequently by apps (see Table 5.1).

Google's tracking was also observed to be deeply integrated into the Android operating system. It has been known that the Google Play Services app – required to access basic Google services, including the Google Play Store – is involved in Google's analytics services [62]. In our network analysis, this app seemed to bundle analytics traffic of other apps and send this information to Google in the background with a time delay. Without access to encrypted network traffic (as explained in Section 5.1), this makes it impossible to attribute network traffic to individual apps from our sample, when such network traffic could also be related to other system apps (some of which, such as the Google Phone app, use Google Analytics tracking themselves). As a consequence, we are likely under-reporting the number of apps that share data with Google, since we only report network traffic that could be clearly attributed.

Consent to tracking is widely absent. Only 9.9% of apps asked the user for consent. Apps that did so contacted a larger number of tracker hosts than those that did not (3.7 with consent vs 2.8 that did not). A slightly larger fraction (12.2% of all apps), informed the user to some extent about their privacy practices; apps in this category also contacted a larger number of trackers than those that did not

(3.6 that informed vs 2.8 that did not). 19.1% of apps that did not ask for consent showed ads, compared to only 2.5% of apps that asked for consent. Once consent was granted, the apps contacted an average of 4.2 tracker hosts (higher than the 3.7 before granting consent, and the 2.8 for apps without any consent flows).

Consent is limited to using or not using an app. Most apps that ask for consent force users into granting it. For instance, 43.7% of apps asking for consent only provided a single choice, e.g. a button entitled 'Accept Policy and Use App' or obligatory checkboxes with no alternative. A further 20.2% of apps allowed users to give or refuse consent, but exited immediately on refusal, thus providing a *Hobson's choice*. Only 42 of the apps that implemented consent (comprising a mere 3.5% of all apps) gave users a genuine choice to refuse consent. However, those apps had some of the highest numbers of tracker hosts, and contacted an average of 5.2 on launch. Among these apps, if consent was granted, the number of tracker hosts contacted increased to 8.1, but, interestingly, an increase was also observed even if data tracking was opted-out (from the pre-consent 5.2 to 7.5 post-opt-out). This increase does not necessarily mean that apps disregard users' consent, but might indicate that apps without consent connect to tracking domains for less invasive types of tracking (e.g. unpersonalised ads instead of personalised ads).

Consent is limited to the personalisation of ads. Consent was often limited to an opt-out from personalised ads. 37 of the 42 apps that implement a genuine choice to refuse consent restrict this choice to limiting personalised advertising; such a choice might make some users wrongly assume that refusing to see personalised ads prevents all tracking (see Figure 5.1 for some common examples). We observed that 23 of these 37 apps (62%; 1.9% overall) used Google's Consent API [183], a toolkit provided by Google for retrieving consent to personalised ads (particularly when multiple ad networks are used). None of the apps using the Google Consent API, however, ended up asking users to agree to further tracking activities, such as analytics. Only 4 apps provided the option to refuse analytics; all 4 of these did so in addition to providing the option to opt-out of personalised advertising. One further app in our sample requested consent to process health data. Since our study,

Tracker	Apps	Expects consent (in EU / UK)	Implements consent (by default)	Mentions consent (in implementation guide)	Discloses local data storage
Google Analytics	50%	Yes	No	No	Yes
Google AdMob	45%	Yes	No	Yes	Yes
Google Crashlytics	29%	Yes	No	No	Yes
Facebook App Events	20%	Yes	No	No	?
Google Tag Manager	19%	Yes	No	No	Yes
Facebook Ads	14%	Yes	Yes*	No	?
Flurry	9%	Yes	No	No	?
Unity Ads	8%	Yes	Yes	No	Yes
Inmobi	8%	Yes	No	Yes	?
Twitter MoPub	6%	Yes	Yes	No	Yes
AppLovin	6%	No	No	No	?
AppsFlyer	5%	?	No	Yes	?
OneSignal	4%	Yes	No	No	Yes

Table 5.2: Consent requirements and implementation for 13 commonly used Android trackers. App shares according to the Exodus Privacy Project [137]. The **trackers in bold** require consent, but do neither implement such by default nor mention the need to do so in their implementation guides. ?: We did not find any information. *: Facebook opts-in users by default to their personalised advertising, unless they disable this behaviour from their Facebook settings or do not use the Facebook app.

Google has deprecated its Consent API and switched to the IAB's Transparency & Consent Framework (see Section 8.3.1 for more details).

5.4 Support and Guidance from Trackers

The previous Section found a widespread absence of consent to third-party tracking in apps. As explained in Section 5.2, both first and third parties have a part to play in facilitating valid consent, and third parties need to take steps to ensure consent obtained by first parties is valid. At the same time, it has been reported that many app developers believe the responsibility of tackling risks related to ad tracking lies with the third-party companies [31], and need clear guidance regarding app privacy [85]. In this Section, we assess the efforts that providers of tracker libraries make to encourage and support app developers in implementing a valid consent mechanism. We focus on the most common libraries in order to understand the current practices across the tracking industry.

5.4.1 Methodology

Our qualitative analysis focuses on the 13 most common tracker companies on Android (according to [137]), and three types of documents that each of them provides: 1) a step-by-step implementation guide, 2) a privacy policy, and 3) further publicly available documentation. While there may be other ways in which providers of tracking libraries support app developers to facilitate valid consent, we reason that these are the standard means by which such support would be provided. Step-by-step implementation guides serve as a primary resource for app developers and summarise the essential steps of implementing a tracker library in code. Since the implementation of consent must be done in code, consent implementation is one essential step for those trackers that require consent.

In assessing this documentation, we assume the perspective of an app developer who is motivated to comply with any explicit requirements mentioned by the tracker provider, and to follow their instructions as to how to do so, but lacks in-depth knowledge about how the GDPR and ePrivacy Directive apply to their use of a given third-party tracking software [87]. We also assume that app developers are likely to read documentation only as far as necessary to make the third-party library functional, often through trial-and-error [184, 185], and stop studying other resources once the tracker implementation is functional, since they are often pressured by time and economic constraints [30, 31, 186].

5.4.2 Results

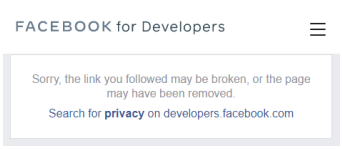
Our results are summarised in Table 5.2. We detail our main findings in the following paragraphs.

Most trackers are unclear about their use of local storage. Whether a tracker accesses and/or stores information on a user's device is essential in determining the need to implement consent, as explained in Section 5.2.2. As such, we would expect to find information stating whether or not access and/or storage takes place as part of the standard operation of the tracker. However, we did not

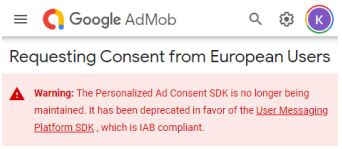
find such information for 6 out of 13 trackers. For the others, this information was difficult to find. AppsFlyer rightly states in its online documentation that 'there are no cookies for mobile apps or devices' [187]. While this is true from a technical perspective, EU and UK law do not differentiate between cookies and other information saved on a user's device. Crucially, we did not find any tracker stating *not* to save information on a user's device. In the absence of such a denial, app developers would run the risk of assuming they do not need to obtain consent for data accessed and/or stored by the tracker.

Most trackers expect app developers to obtain consent. Despite being unclear about their use of local storage, a closer inspection of the tracker policies and documentation found that most trackers instruct developers to request consent from EU users (11 out of 13). AppLovin is an exception, but does require consent if developers want to show personalised ads (which tend to be more lucrative than contextual ads). For AppsFlyer, we could not find any information regarding the need to ask users for consent. The need to ask for consent was sometimes difficult to find, and required a careful reading of the policies and documentation provided. Some developers are bound to overlook this, and unnecessarily compromise on the users' right to choose over tracking.

Few trackers implement consent by default. We further inspected whether tracker libraries provide their own consent implementation. If they do, an app developer would not need to make any further modifications to the app code. However, only a minority of tracker libraries (3 out of 13) integrate an implementation of user consent by default, and none of the five most common trackers do so. Unity Ads and Twitter MoPub provide consent flows that are automatically shown, without further action by the app developer. Facebook Ads only shows ads, if the app user 1) has agreed to personalised ads in their Facebook account settings, and 2) uses the Facebook app on their phone. However, Facebook opts-in users by default to their personalised advertising, unless they disable this behaviour from their Facebook settings (checked 14 February 2021). While Google AdMob provides a consent library, this is not implemented by default. Indeed,



(a) Facebook links to a page that is supposed to explain how to implement consent in practice.



(b) Google AdMob links to an outdated library, creating unnecessary friction for consent implementation.

(c) Flurry links to a broken GDPR guide.

Figure 5.2: Many trackers provide information on what developers need to know to implement consent. These guides are often difficult to find, hard to read, and poorly maintained. 3 out of 13 common trackers linked to unmaintained or broken pages.

Google AdMob expects the app developer to retrieve consent from the user, but shows personalised ads even if the developer does not implement their consent library.

Limited disclosure of consent requirements in step-by-step guides. We find that 3 out of 13 tracker libraries disclose the potential need for consent in their step-by-step implementation guides. This is despite 11 out of 13 trackers mentioning the need to implement consent in other places of their online documentation. Google AdMob mentions the need to retrieve consent among other 'examples of actions that might be needed prior to initialization [of AdMob]' [188]. Inmobi points out that developers need to 'obtain appropriate consent from the user before making ad requests to InMobi for Europe' [189] in the Section on 'Initializing the SDK'. AppsFlyer offers developers to 'postpone start [of the tracker library] until you receive user consent due to GDPR or CCPA requirements, etc.' [190] in Section 3.4 on 'Delay SDK initialization'. It is not clear from these three implementation guides what other reasons are to 'delay initialisation' beyond legal compliance, and why this is not clarified. At least 6 out of 13 trackers require consent, but neither implement such by default nor inform app developers of the need to do so in the implementation guides. If AppLovin needs consent (despite not stating to do so, but as suggested by our legal analysis in Section 5.2), this figure would increase to 7 out of 13 trackers.

Compliance guidance: often provided, but sometimes difficult to find, hard to read, and poorly maintained. Many tracker companies provide additional information on GDPR compliance and consent implementation on a separate website as part of their online documentation. We found some compliance

guidance (with varying levels of detail) for all trackers except the Google Tag Manager. Excluding the 3 trackers implementing consent by default, a developer needs an average of 1.56 clicks to reach these compliance guides. For AppLovin, a developer must click 'Help Center', then 'Getting started & FAQ', and lastly 'User opt-in/opt-out in the AppsFlyer SDK'. Facebook required developers to click 'Best Practices Guide' and then 'GDPR Compliance guide'. While this GDPR compliance guide provides some guidance on the implementation of consent, the link to Facebook's 'consent guide' with practical examples of how to implement consent was broken. Also, the framing as 'Best Practices' suggests the optionality of legal compliance. For OneSignal, developers must first click 'Data and Security Questions' and then 'Handling Personal Data'.

The compliance guides (excluding code fragments) reached a mean Flesch readability score [191] of 41.8, as compared to 50.6 for the step-by-step implementation guides (where 100 means 'very easy', and 0 'very difficult' to read). Both the implementation and compliance guides are 'difficult' to read, with the compliance guides somewhat more so. For 3 of the 13 trackers, we were directed to broken or outdated links (see Figure 5.2). Google AdMob linked to an outdated consent strategy, while the Facebook SDK and Flurry linked to non-existing pages (returning `404` errors). We found other pages with compliance information for each of these trackers, but broken guidance can act as a deterrent for developers who want to implement consent and follow their legal obligations. However, since this Chapter was published, the broken links in the documentation of the Flurry and Facebook trackers were fixed.

5.5 Discussion

Consent is an integral part of data protection and privacy legislation, both in the EU and the UK, and elsewhere. It is all the more so in the context of third-party tracking, for which consent appears to be typically a precondition, as analysed in Section 5.2. Not only has this been emphasised by multiple DPAs, but is also

acknowledged by tracking companies themselves in the documentation they make available to app developers. Relying on the controller's legitimate interests – the only conceivable alternative to consent under EU and UK data protection law – would likely fall short of passing the balancing test outlined in Article 6(1)f GDPR. This also follows from the requirement to obtain consent prior to storing or accessing information on a user's device, under the ePrivacy Directive.

Against this backdrop, we analysed 1,297 mobile apps from Google Play in Section 5.3 and discovered a widespread lack of appropriate mechanisms to gather consent as required under the applicable regulatory framework. We found that, while the guidelines of many commonly used tracker libraries require consent from EU and UK users, most apps on the Google Play Store that include third-party tracking features do not implement any type of consent mechanism. The few apps that require data subjects to consent do so with regard to personalised advertising, but rarely for analytics – despite this being one of the most common tracking practices. Where an opt-out from personalised advertising was possible, the number of tracker domains contacted decreased only slightly after opting-out, hinting at continued data collection when serving contextual advertising. These observations are at odds with the role of consent as the only viable option to justify the processing of personal data inherent in third-party tracking.

As detailed in Section 5.3, the fact that only 9.9% of the investigated apps request any form of consent already suggests widespread violations of current EU and UK privacy law. This is even before considering the validity of the consent mechanisms put in place by that small fraction of apps. As underlined in Section 5.2, consent must be 'freely given', 'informed', 'specific' and 'unambiguous'. The findings outlined in Section 5.3 suggest that most apps that do implement consent force users to grant consent, therefore ruling out its qualification as 'freely given'. The same goes for the 43.7% of those apps that do not provide data subjects with the possibility to consent separately for each purpose, but instead rely on *bulk* consent for a wide array of purposes.

When considering both the absence of any form of consent in more than 90% of the investigated apps and the shortcomings inherent in the few consent mechanisms that are implemented by the remaining sample, we infer that many mobile apps fail short of meeting the requirements stemming from EU and UK data protection law. Our analysis does not even consider the fact that consent is only one of a variety of legal rules that third-party tracking needs to comply with. Breaches of other legal principles – such as data minimisation, purpose and storage limitation, security and transparency – might be less visible than a lack of consent and harder to analyse, but no less consequential.

We further found that one of the reasons for the lack of consent implementation in apps might be inadequate support by tracker companies [31, 85]. Studying the online documentation of the 13 most commonly used tracker libraries in Section 5.4, only 3 trackers implemented consent by default, and another 3 disclosed the need to implement consent as part of step-by-step implementation guides. These step-by-step guides serve as a primary resource for app developers, and can give a false impression of completeness when in fact additional code needs to be added for many trackers to retrieve user consent. This is true for at least 6 out of 13 trackers, including Google Analytics and the Facebook App Events SDK, which likely need consent, but neither disclose this in their implementation guides nor implement such consent by default. While most trackers provide some compliance guidance, we found that this can be difficult to find, hard to read, and poorly maintained. Whatever the reasons for the lack of consent, the result is an absence of end-user controls for third-party tracking in practice.

Lastly, it is worth highlighting that Google, which is both the largest tracking company and the main developer of Android, faces conflicts of interest with respect to protecting user privacy in its Google Play ecosystem [82, 192, 193]. The company generates most of its revenue from personalised advertising, and relies on tracking individuals at scale. Certain design choices by Google, including its ban of anti-tracking apps from the Play Store, its recent action against modified versions of Android, and the absence of user choice over AdID access for analytics on Android

(as opposed to iOS), create friction for individuals who want to reduce data collection for tracking purposes, and lead to increased collection of personal data, some of which is unlawful as our legal analysis has shown.

5.6 Limitations

It is important to acknowledge some limitations of our methodology. Our analysis in Section 5.3 used dynamic analysis, and not all tracking might be detected. We only inspected network traffic before and shortly after consent was given. Apps might therefore conduct more tracking during prolonged app use. Besides, we only reported the network traffic that could be clearly attributed to one of the apps we studied, potentially leading to under-reporting of the extent of Google's tracking (as explained in Section 5.3).

While the reported tracking domains can be used for tracking, they might also be used for other non-tracking purposes. However, it is the choice of the tracking company to designate domains for tracking. Indeed, there is an incentive for tracking companies to bundle different purposes under one domain so as to make blocking on non-essential network traffic more difficult for DNS-based ad blockers, but thereby also conflicting with the GDPR's transparency principle (Article 6(1) GDPR).

We do not study the contents of network traffic because apps increasingly use certificate pinning (about 50% of the studied apps used certificate pinning for some of their network communications). As for our second study in Section 5.4, we studied the online documentation of tracker libraries with great care, but did not always find all relevant information, particularly regarding the local storage of data on a user's device. Where this was the case, we disclosed it (e.g. see Table 5.2).

5.7 Conclusions & Future Work

This Chapter analysed the legal requirements for consent to tracking in apps, and found an absence of such consent in practice based on an analysis of a representative sample of Google Play apps. This, in turn, suggests widespread violations of EU and

UK data protection law. Simple changes by software intermediaries (such as Google and Facebook), including default consent implementations in tracker libraries, better legal guidance for app developers, and better privacy options for end-users, could improve the status quo around app privacy significantly. However, it is doubtful that these changes will happen without further intervention by independent parties – not only end-users, but also policymakers and regulators – due to inherent conflicts between user privacy and surveillance capitalism.

While the web has seen a proliferation of deceptive and arguably meaningless consent banners in recent years [5, 7], we should hope that mobile apps will not see similar mass adoption. Rather, we aim to influence the current policy discourse around user choice over tracking and ultimately to make such choice more meaningful. As Apple has demonstrated with iOS 14, system-level user choices can standardise the process of retrieving user consent and make hidden data collection, such as tracking, more transparent to end-users. We will explore the impact of these changes in Chapter 7.

Future work. An overarching question for future work is the extent of the legal obligations faced by the many actors involved in the third-party tracking ecosystem, ranging from app developers to providers of tracker libraries and mobile operating systems. This is inextricably linked to their qualification as 'controllers', a legal notion whose boundaries still remain controversial, despite recent jurisprudence [117, 194, 195] and detailed guidance [196, 197]. Our analysis highlighted how simple changes in the software design can have significant effects for user privacy.

Moreover, while the US – unlike many developed countries – lack a federal data protection law, there exists a variety of specific privacy laws, such as COPPA to protect children and HIPAA to protect health data, as well as state-level privacy laws, including CCPA in California. Some of these laws foresee consent requirements similar to EU and UK law. We leave it to further work to assess how widely apps comply with the consent requirements of US privacy legislation.

6

Choice between iOS and Android

We explored in Section 2.3.2 how Google and Apple govern their respective ecosystems with fundamentally different strategies. Despite this, previous research on smartphone privacy has mainly focused on one of these companies Google and the Android ecosystem [2, 47, 59, 67–74, 76]. Limited research exists on Apple's iOS and App Store ecosystem [4, 58], which has a market share of nearly two-thirds in the US [198, 199]. One reason for the limited existence of iOS privacy research has been the lack of publicly available analysis tools, paired with the encryption of iOS apps and the uncertain legality of their decryption (we discuss this more in Section 6.1.1). Knowledge of the app ecosystem is important so that consumer choice between platforms can be informed on privacy grounds, but moreover for effective regulation [19, 48, 200–202] and democratic debate regarding these increasingly important pieces of digital infrastructure.

Given the differences between these business models and the greater emphasis on privacy by Apple, it would be reasonable to assume that the iOS ecosystem would be more privacy-protective in general, in terms of the kinds of data that can be shared, and the extent of third-party sharing. However, little recent empirical research has tested these assumptions in detail, by comparing the privacy practices of apps on the two ecosystems at scale. This Chapter fills this gap, by examining

the privacy behaviours of apps on the Apple App Store and Google Play, comparing them explicitly, and examining how particular design decisions underlying the two ecosystems might affect user privacy.

Empirical contributions. Given that there are multiple dimensions of privacy, and a corresponding multiplicity of ways to measure it, we adopt a mixture of different indicators and scales to examine each ecosystem along several complementary facets, as follows:

1. *Code Analysis* of a representative sample of 12k apps from each platform to assess commonly studied privacy metrics (e.g. permissions and presence of tracking libraries) at scale and across platforms.
2. *Network Traffic Analysis* of the same 12k apps from each platform to study apps' real-world behaviour.
3. *Company Resolution* to reveal the companies ultimately behind tracking, as well as the jurisdictions within which they reside.

Using the privacy footprints built from our analyses, we find and discuss violations of privacy law and limited compliance with app stores' data collection policies. We note that while there exist a few other studies that have looked at *security vulnerabilities* in larger numbers of iOS apps [77, 134, 203], this present study is the largest study of *privacy aspects* of apps across Android and iOS to date and of privacy in iOS apps since 2013 [33]. Analysing apps last updated 2018–2020, we study app privacy shortly before Apple's introduction of mandatory opt-ins to tracking in 2021 with iOS 14.5.

Technical contributions. We present a methodology for large-scale and automatic download, privacy analysis, and comparison of apps from the Google Play and Apple App Stores. So far, no comparable tools have existed in the public domain, despite such tools being necessary to understand app privacy at large, and to hold the platform gatekeepers to account. Compared to previous analysis tools for iOS, our approach does not rely on the decryption of apps. We make our tools and dataset, including the raw app data, publicly available at `https://platformcontrol.org/`.

	Android Only		iOS Only		Android & iOS		
	Viennot [76]	Binns [4]	Agarwal [33]	Egele [75]	Han [74]	Ren [71]	This Chapter
Publication Year	2014	2018	2013	2011	2013	2016	2021
Total Apps	1m	1m	226k	1.4k	2.6k	200	24k
Static Analysis	✓	✓	x	✓	✓	x	✓
Dynamic Analysis	x	x	✓	x	x	✓	✓
Tracking Libraries	✓	✓	x	x	✓	x	✓
Permissions	x	x	x	x	✓	x	✓
PII Usage	x	x	✓	✓	x	✓	✓

Table 6.1: Previous papers studying *privacy* properties of iOS and Android apps. We only include a small subset of important 'Android Only' studies. We do not include papers that focus on security vulnerabilities of apps.

Structure. The remainder of this Chapter is structured as follows. We first summarise the challenges in analysing iOS apps and review related work in Section 6.1. Next, we introduce our app analysis methodology of 12k apps from each app platform in Section 6.2. We then turn to our results from the code and network traffic analyses in Section 6.3, with a focus on compliance of apps with privacy law. We discuss limitations in Section 6.4, and conclude the Chapter and outline directions for future work in Section 6.5.

6.1 Background

6.1.1 Challenges on iOS

While many studies have analysed privacy in the Android ecosystem, comparatively much less is known about iOS. One reason for this lies in the few unique challenges that the Apple ecosystem poses. First, the closed-source nature of the underlying operating system (iOS), including the use of Apple-only programming tools (Xcode), languages (Objective-C and Swift) and compilers, complicates analysis efforts. Previous work managed to decompile a subset of iOS apps, but no universal decompilation tools exist [75, 204]. Another challenge is Apple's *FairPlay DRM*, which makes accessing and analysing apps' code more difficult than on Android. Decryption is possible, but relies on access to a physical device and takes time [75, 77, 134]. Depending on the jurisdiction, there might also be legal challenges related to the decryption of iOS apps, since this circumvents copyright protections (though

arguably not particularly effective ones). However, there exist exemptions for research purposes in some jurisdictions (e.g. the UK and US), the analysis of which is beyond the scope of this dissertation. By contrast, Google only encrypts *paid* apps (not free ones) when downloaded from its Play Store.

Apple's use of proprietary technologies and copyright protections acts as a deterrent to developing scalable download and privacy analysis tools for iOS. No publicly available, scalable tools exist for the Apple App Store (unlike for Google Play) [4, 58, 134]. However, such tools are necessary to understand the iOS ecosystem at large, and to hold the platform gatekeepers to account. This Chapter addresses this gap by introducing tools and methods for scalable analysis of iOS apps without relying on app decryption (see Section 6.2.2.1). This allows us to share our tools publicly, without having to worry about uncertain liability.

6.1.2 Research Gap

Previous research extensively studied privacy in mobile apps. As discussed in Section 2.2.4, there exist two main methods in the academic literature for such studies: dynamic and static analysis. Key pieces of literature are summarised in Table 6.1, and are discussed next. Table 6.1 evaluates prior work based on the analysis technique used (static or dynamic) and on the privacy properties studied: (i) *tracking libraries*, (ii) *permissions*, and (iii) *PII usage*.

Tracking libraries. Several studies exist that examine the presence of tracking libraries in apps. For instance, Viennot et al. [76] analysed more than 1 million apps from the Google Play Store in 2014, and found that 36% of analysed apps contained the Google Ads library, 12% the Facebook SDK, and 10% Google Analytics. Similarly, Binns et al. [4] decompiled and analysed third-party data collection in about 1m Google Play apps in 2018. The authors found a strong concentration of data collection with very few tracker companies ('trackers'), with Google and Facebook being the most prominent. Chen et al. [77] decompiled ~1.3m Android and ~140k iOS apps, and found potentially malicious libraries in 7% of Android and 3% of iOS apps in 2016.

Permissions. Analysing permission use by apps has a long history in app research [2, 74, 102, 205–211]. For instance, Han et al. [74] decompiled and analysed 1,300 pairs of iOS and Android apps in 2013. They found that iOS apps accessed sensitive data significantly more often than their Android counterparts. Advertising and analytics libraries accounted for a third of these accesses. The analysis of permissions only gives a partial picture of apps' privacy practices, since apps tend to request more permissions than necessary [205], but may never access the information associated with the permission. Moreover, some Android apps have even been found circumventing the permissions system [102].

PII usage. There are several approaches to study PII usage in apps. Some approaches, such as the one taken by Agarwal and Hall [33] in 2013, examine *access* to sensitive data by intercepting API calls in a jailbroken iOS device. Since access does not always lead to *transmission*, recent work has shifted to a network-based approach to detect PII *exposure* over the network [2, 67, 69–71, 212]. For example, Ren et al. [71] developed a VPN server to detect the sharing of PII independent of the mobile operating system in 2016.

Our work. In this Chapter, we provide an updated study of privacy practices in apps across Android and iOS at a sufficient scale. Most of the studies discussed above examine either the Android or the iOS ecosystem. The number of comparative studies is limited, so we seek to address this gap. Unlike previous work, we analyse iOS apps without relying on app decryption or only traffic analysis, to produce rich insights about app privacy at scale through both dynamic and static analysis, and to make our analysis toolchain public at `https://platformcontrol.org/` without having to worry about uncertain liability.

6.2 Methodology

In this Section, we describe our analysis methodology, depicted in Figure 6.1. We begin by detailing our app selection and download process in Section 6.2.1. Next, in Section 6.2.2, we present our method for code analysis, which allows us to extract

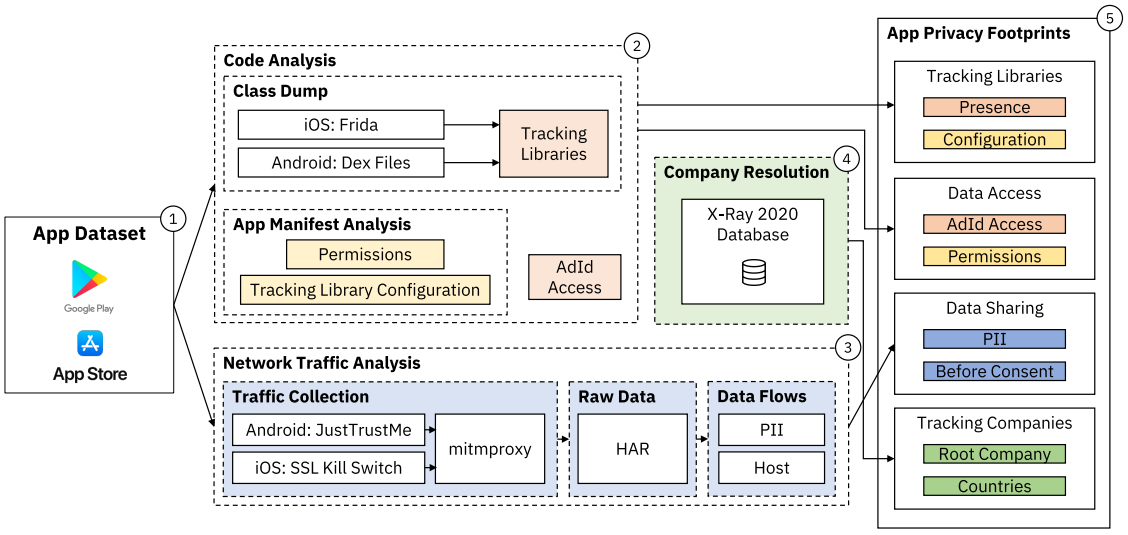


Figure 6.1: Overview of our analysis methodology (Section 6.2): First, (1) we select and download 12k apps from the Google Play and Apple App Stores each (Section 6.2.1). We then perform a (2) **Code Analysis** (Section 6.2.2): (i) we inspect the list of class names (obtained from `*.dex` files on Android and Frida class dumps on iOS) for known tracking libraries; (ii) we check if apps can access the AdId; and (iii) we analyse the App Manifest to obtain a list of permissions and also to determine the privacy configuration of popular tracking libraries. Third, (3) we conduct a **Network Traffic Analysis** (Section 6.2.3): we disable certificate validation and execute each downloaded app while using `mitmproxy` to capture network traffic in the `HAR` format. We analyse the captured traffic for occurrences of PII and contacted host names. Finally, (4) we perform a **Company Resolution** (Section 6.2.4) to obtain a list of companies behind tracking, their owner companies, and the countries of these companies. We use the X-Ray 2020 database for this analysis and resolve the companies behind both the identified tracking libraries in (2) and the contacted hosts in (3). The results of this analysis (Section 6.3) are detailed **App Privacy Footprints** (5) of the downloaded apps, that allow for comparison of privacy characteristics between the two platforms.

the following information about each app (without the need to decrypt iOS apps): what tracking libraries are used, how they are configured, which permissions are requested, and whether or not the AdId is accessed. Afterwards, in Section 6.2.3, we describe how we collected and decrypted apps' network traffic and analysed it for PII exposure. Finally, in Section 6.2.4, we provide details on how we resolved tracking activities (found by both code and network analysis) to the companies behind them and their country of origin.

6.2.1 App Dataset and Selection

As the foundation for the analysis in this Chapter, we use the app dataset that we collected in 2020 in Section 3.1.1. We only considered apps that were released or updated in 2018 or later to focus on apps that are still in use. Due to the size of the dataset (containing almost 560k apps released or updated in 2018 or later), we selected a random subset of apps of 24,000 apps (12,000 from each platform) for further analysis in this Chapter.

Statistical extrapolation from our sample. In this Chapter, we are interested in studying trackers, and thus we need to ensure that the results we gather on tracking activities in our app corpus can be extrapolated. Here, we will argue that our corpus of 24,000 apps is statistically sound when it comes to tracking libraries. For a description of how we identify tracking libraries, see Section 6.2.2. We chose to download more than 10,000 apps for each platform to bring down the margin of the 95% confidence interval (and thereby, the sampling error in our dataset) for the tracker prevalence $\overline{X_T}$ to less than 2%, for every studied tracker T, assuming an underlying normal distribution due to the law of large numbers:

$$\overline{X_T} \sim N(\mu, \sigma^2)$$

Studying a random subset of 100 or even 1000 Android apps would not suffice to reach this sampling error margin of 2%. For example, the expected 95% confidence interval for containing the Facebook SDK was (19.2%, 37.0%) for a sample of 100 apps from our dataset, yielding an expected sampling error margin of 17.8%. For a sample of $1,000$ Android apps: (25.3%, 30.9%), yielding an expected sampling error margin of 5.6%. However, in our dataset of $12,000$ Android apps, 28.1% of apps contained the Facebook SDK; the 95% confidence interval was (27.3%, 28.9%). In conclusion, while we focus on a subset of the overall apps, our results can be extrapolated to the larger dataset, and across all apps on the app stores updated since 2018, with limited error.

Additionally, where appropriate, we conduct permutation tests (using 10,000 permutations) to assess the statistical significance of any quantitative comparisons.

In our tests, we use the difference in mean as our test statistic. Where we do not find statistical significance ($p > 0.05$), we also report 95% confidence intervals.

Identification of cross-platform apps. While the majority of this Chapter analyses the set of 24,000 downloaded apps, Section 6.3.5 examines *cross-platform apps* – using a simple similarity algorithm that examined terms from both app titles and app identifiers as follows: We first tokenised, counted and frequency-weighted terms from app titles and app identifiers for all 560k iOS and Android apps (i.e. all apps from our dataset in Chapter 3 that were released or updated in 2018 or later) using TF-IDF, then computed cosine similarities between pairs of the resulting vectors. Among the 24k downloaded apps, we considered only those apps as cross-platform that had a cosine similarity of at least 95%. This amounted to 13.7% of downloaded Android apps, and 12.8% of iOS apps.

6.2.2 Code Analysis

In this section, we describe how we analyse the apps' code in order to assess the usage and configuration of tracking libraries, access to the AdId, and requested permissions (see step 2 in Figure 6.1.

6.2.2.1 Tracking Libraries: Presence

Tracking library detection. We first obtained the class names in Android apps directly from their corresponding `*.dex` files, while for iOS, we used the `Frida` dynamic instrumentation toolkit to dump class names from apps. Note that decryption of iOS apps was not necessary with this `Frida`-based approach. We then studied what class names occurred in at least 1% of Android or iOS apps and are related to tracking, similar to Han et al. [74]. We resolved class names to tracking libraries using various online resources, including the Exodus Privacy project for Android apps [137] and the CocoaPods Master repository for iOS ones [213] (containing information on class signatures) as well as trackers' online resources (documentation and GitHub repositories). We identified a total of 40 tracking libraries of interest, all of which existed for both Android and iOS,

except for Google's Play Services (which was present on Android only) and Apple's SKAdNetwork (which was present on iOS only).

Impact of obfuscation. While some very popular apps may use code obfuscation to hide their tracking activities intentionally, we found that it had little effect on our overall analysis. By default, iOS apps do not apply any obfuscation to the class names, and developers are well known to be subject to a 'default bias' (i.e. not to change default settings) in the literature [31, 53, 86, 87]. As for Android, we found similar results by checking against the obfuscation-resilient LibRadar++ library [59, 153]. An important reason for this result is that, while tracking libraries may obfuscate their internal code, obfuscating user-facing APIs is difficult [86]. Further, many tracking libraries use inter-app communication and cannot easily obfuscate their communication endpoints [102]. We do not use LibRadar++ for our overall analysis, since it is closed-source, no longer maintained, has an outdated database of library signatures (last updated in 2018), and struggled with different library configurations (for instance, Google Firebase is a set of different libraries, including advertising and analytics components that share some of the same code, but LibRadar++ summarised all these components as `com.google.firebase`). We also tried LibScout [214, 215] for library detection, but found that it also missed essential libraries and took a long time to execute.

6.2.2.2 AdId Access

The AdId is a unique identifier that exists on both iOS and Android. It allows advertisers to show more relevant ads for users (e.g. by avoiding showing the same advert twice in two different apps), but it can also be used to create fine-grained profiles about app users. This is something many users may not expect and that can lead to potential violations of data protection law, including the need to seek consent before tracking (see Chapter 5). The AdId is also the only cross-app identifier that may be used for advertising on Android, but might additionally be used for analytics [216]. That is why AdId access might be an upper bound on the use of any form of analytics on Android (including personalised ads); there are no incentives not to use

the AdId for these purposes. While users can theoretically reset the AdId, most do not know how or why to do so [43, 53]. Starting with iOS 14.5 in 2021, the operating system has switched from an opt-out to an opt-in mechanism for apps' use of the AdId; in our study, we will assess privacy in the app ecosystem immediately before this policy change. We detected potential access to the AdId by checking for the presence of the `AdSupport` class and the system interface `IAdvertisingIdService` in the app code on iOS and Android, respectively. After our study, Google has begun to roll out a new *opt-out* framework from the AdId, in response to Apple's new *opt-in* App Tracking Transparency framework that we will discuss in the next Chapter.

6.2.2.3 App Manifest Analysis

Permissions. Permissions form an important part of the security models of Android and iOS as they protect sensitive information on the device. We extract the permissions used by the apps in our dataset by parsing the app manifest files. At the time of data collection, Android defined a total of 167 permissions, 30 of which were designated as *dangerous permissions* by Google and require user opt-in at run-time. Similarly, Apple defined 22 permissions that needed to be disclosed in the app manifest. All of these require user opt-in. We only include permissions defined by the Android or iOS operating system, and exclude custom permissions by other vendors (used by some Android apps). While the targeted OS version can affect what permissions apps can request, only a few new permissions have been added in 2018–2020 and we did not consider this aspect further; in our results, none of the top 10 permissions on either platform has been added in the period 2018–2020, so this should not significantly affect our reported descriptive statistics.

In addition to reporting statistics on general permissions usage, we further focus on the ones that both Apple and Google agree to be particularly dangerous and need user opt-in. There are a total of 7 such *cross-platform permissions* that exist on both platforms: Bluetooth, Calendar, Camera, Contacts, Location, Microphone, and Motion. This total number is small compared to the overall number, since we excluded and summarised some permissions to overcome the

different functionality and granularity in permissions across the platforms. For instance, Android discriminates between read and write permissions for contacts and calendar, but we have summarised them as *Contacts* and *Calendar*, respectively.

Tracking library configuration. Many tracking libraries allow developers to restrict data collection using settings in the app manifest, e.g. to disable the collection of unique identifiers or the automatic SDK initialisation at the first app start. This can help setting up tracking libraries in a legally compliant manner. For the Facebook SDK, these options were only added after public backlash over the mandatory and automatic sharing of personal data at the first app start, potentially violating EU and UK data protection law [217]. We focus on the privacy settings provided by some of the most prominent tracking libraries: Google AdMob, Facebook, and Google Firebase.

6.2.3 Network Traffic Analysis

In this Section, we discuss our network traffic analysis process (step 3 in Figure 6.1).

App execution and network traffic capture. We opened every app automatically on a real device – a Google Nexus 5 running Android 7 and an iPhone SE 1st Gen with iOS 14.2 – for 30 seconds without user interaction. We captured apps' network traffic using `mitmdump` to study apps' data sharing with tracking domains. Tracking libraries are usually initialised at the first app start and often without user consent [6, 70, 178], which we aimed to detect with this approach. We did not perform any further automated actions on the studied apps, since there did not exist established approaches (like the UI Exerciser Monkey on Android or more sophisticated approaches [218]) to instrument *arbitrary* iOS apps.

Device configuration. We disabled certificate validation using `JustTrustMe` on Android and `SSL Kill Switch 2` on iOS, after gaining system-level access on both devices (known as 'root' on Android and 'jailbreak' on iOS), in order to read apps' HTTPS traffic. We would have liked to use a more recent version of Android, but we found that disabling certificate validation was unstable on the latest versions of Android. We note that several identifiers are inaccessible as of Android 10, but

apps should behave similarly otherwise. We uninstalled or deactivated pre-installed apps, and were not logged into an Apple or Google account. On both phones, we did not opt-out from ad personalisation from the system settings, thereby assuming implicit user opt-in to apps' use of the AdId.

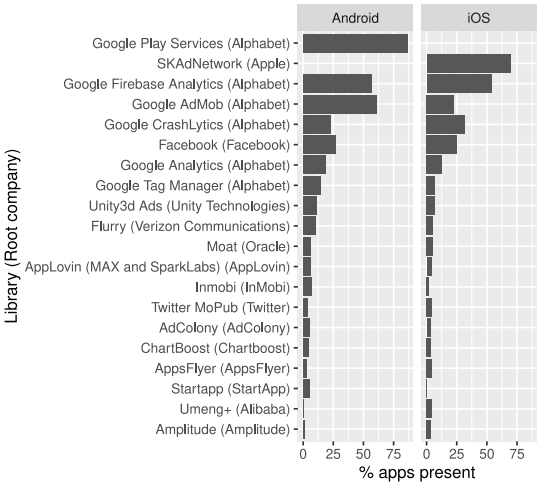
PII analysis. To analyse the sharing of PII and other personal data, we conducted a case-insensitive search on the network traffic for the following identifiers as well as common transformations thereof (MD5, SHA-1, SHA-256, SHA-512, URL-Encoding): Advertising ID, Android Serial Number, Android ID, phone and model name, and Wi-Fi MAC Address. We have refrained from analysing further PII, such as location, contacts or calendar, due to a lack of instrumentation methods for iOS to get around opt-in permission requests. We also assembled a list of contacted host names.

6.2.4 Company Resolution

As part of our following analysis, we also explore which companies are ultimately behind tracking, and in which jurisdiction these are based (step 4 in Figure 6.1). We combine the insights from both the studied tracking libraries (Section 6.2.2.1), as well as all tracking domains observed in at least 0.5% of apps' network traffic (Section 6.2.3). For this purpose, we use the X-Ray 2020 database that we introduced in Section 3.1.2 and was updated such that it includes all relevant tracking libraries and domains – on iOS and Android.

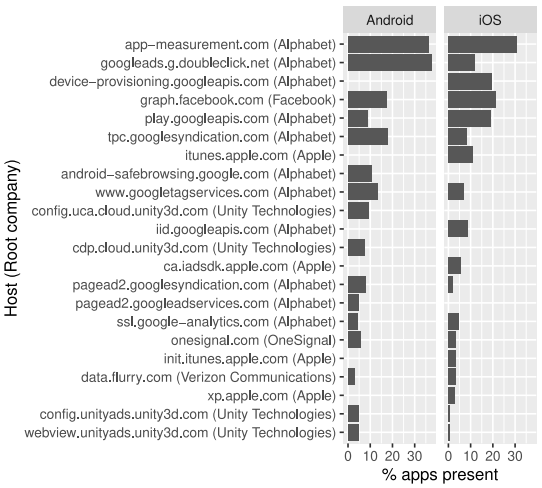
6.3 Results

In this Section, we present our findings from analysing 24,000 apps from iOS and Android (step 5 in Figure 6.1). We analysed 0.86 TB of downloaded apps, and collected 24.2 GB of data in apps' network traffic. Installing and instrumentation failed for 124 Android and 36 iOS apps; we have excluded these apps from our subsequent analysis.



	Median	Mean	Q1	Q3	Count > 10	None
Android	3	3.8	2	5	3.73%	11.27%
iOS	3	3.1	1	4	3.13%	20.65%

(a) Top tracking libraries in app code.



	Median	Mean	Q1	Q3	Count > 10	None
Android	2	2.7	1	4	1.99%	18.56%
iOS	2	2.4	0	4	1.35%	31.54%

(b) Top tracking hosts contacted at the first app start.

Figure 6.2: Third-party libraries and contacted tracking domains of apps, as well as the companies owning them (in brackets). Shown are the top 15 tracking libraries and domains from each platform.

First, we focus on the tracking libraries found from the code analysis and whether or not they were configured for data minimisation (Section 6.3.1). Next, in Section 6.3.2, we analyse potential data access of apps, by examining their permissions and their access to the AdId. Following up, in Section 6.3.3 we report on the actual data sharing of apps before consent is provided, as well as the observed exposure of PII in network traffic. Afterwards, we explore the complex network of companies behind tracking and their jurisdictions (Section 6.3.4). Lastly, we focus on cross-platform (Section 6.3.5) and children's apps (Section 6.3.6). Cross-platform apps have received attention in previous studies, but might feature different privacy properties than apps in the ecosystem overall. Children's apps must adhere to stricter privacy rules, arising both from legal requirements (e.g. COPPA [163] in the US and GDPR in the EU) and the policies of the app store providers.

6.3.1 Tracking Libraries

6.3.1.1 Presence

Apps from both platforms widely use tracking libraries (see Figure 6.2a). The median number of tracking libraries included in an app was 3 on both platforms. 3.73% of Android apps contained more than 10, compared to 3.13% on iOS. 88.73% contained at least one on Android, and 79.35% on iOS.

The most prominent tracking library on Android is the Google Play Services (in 87.3.% of apps) – a technology that is ultimately owned by Google's parent company Alphabet. This library provides essential services on Android devices, but is also used for advertising and analytics purposes. The most prominent library on iOS is the SKAdNetwork library (in 69.6% of apps). While part of Apple's privacy-preserving advertising attribution system, this library discloses information about what ads a user clicked on to Apple, from which Apple could (theoretically) build user profiles for its own advertising system. Google's advertising library ('AdMob') ranks second on Android, and occurs in 61.7% of apps. One factor driving the adoption of this library on Android might be that it not only helps developers show ads, but also provides easy access to the AdId (although developers could also implement this manually). However, this dual use of the tracking library might increase Google's reach over the mobile advertising system, by incentivising the use of AdMob. Google Firebase is the second most popular tracking library on iOS, occurring in 53.9% of apps, as compared to 57.6% on Android. Facebook (nowadays called 'Meta'), the second largest tracker company, has a far smaller reach than Google, and is only present in 28.0% of apps on Android and 25.5% on iOS. Few tracking services are more popular on iOS: Google Crashlytics occurred in 31.8% of apps, and 23.8% on Android. MoPub, a Twitter-owned advertising service, was present in 4.71% of iOS apps, and 4.25% on Android. Overall, tracking services are widespread in both ecosystems, but slightly more so on Android, likely in part due to Google's dual role as a dominant advertising company and platform

gatekeeper on Android. However, Google also has a significant presence on iOS, highlighting its dominance in both smartphone ecosystems.

We note that certain libraries have sub-components which can be loaded individually and have different consequences for privacy. For instance, both Google Play Services and Google Firebase bundle a range of different services, from which developers can choose. Further, certain libraries provide configuration options that also affect privacy. While we do not consider all the sub-components of libraries in this study, we do analyse the libraries' configurations, as discussed in the next Section.

6.3.1.2 Configuration for Data Minimisation

Only a small fraction of apps made use of data-minimising SDK settings in their manifest files, e.g. to retrieve user consent before sharing data with trackers. At the same time, 'data minimisation' is one of the key principles of GDPR, as laid out in Article 5, and user opt-in is usually required prior to app tracking in the EU and UK (see Chapter 5). However, we found that the vast majority of developers did not change trackers' *default options* which might lead to more data sharing than necessary.

Among the apps that used Google AdMob, 2.2% of apps on iOS and 0.8% on Android chose to delay data collection. Among the apps using the Facebook SDK, less than 5% (2.3% on Android, 4.6% on iOS) had delayed the sending of app events, less than 1% (0.4% on Android, 0.9% on iOS) had delayed the SDK initialisation, and less than 4% had disabled the collection of the AdId (0.9% on Android, 3.0% on iOS). Among apps using Google Firebase, 0.5% had permanently deactivated analytics on Android and 0.4% on iOS, 1.2% had disabled the collection of the AdId on Android and 0.1% on iOS, and 1.2% had delayed the Firebase data collection on Android and 0.5% on iOS.

6.3.2 Data Access

6.3.2.1 AdId Access

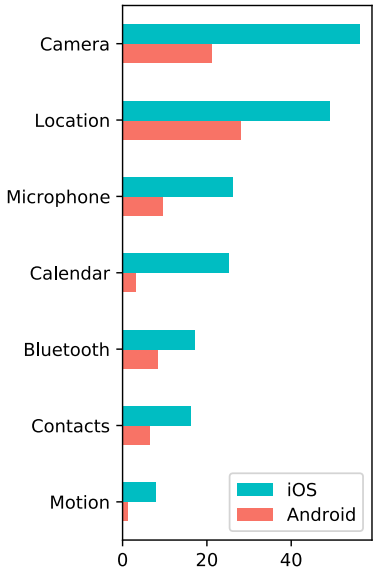
Potential access to the AdId was more widespread among Android apps than iOS ones. Among the studied apps, 86.1% of Android apps could access the AdId, 42.7% on iOS, allowing them to track individuals across apps easily.

Advertising and AdId access were often linked. Of those apps with Google AdMob, 100% on iOS and 99.6% on Android had access to the AdId. We had similar results for the next most popular advertising services: of apps with Unity3d Ads, more than 99% of apps accessed the AdID; similarly for Moat (100% of apps), and Inmobi (more than 94% of apps). Conversely, about 71.3% of Android apps and 53.4% of iOS apps with AdId access used Google AdMob, but less than 20% used Unity3d Ads, Moat or Inmobi. If we assume, for argument's sake, that an app shows personalised ads if and only if it has AdId access (because there is hardly any reason for apps not to use the AdId for personalised ads), this suggests that Google AdMob was present in the majority of apps with personalised ads. This points to a high *market concentration* towards Google in the digital advertising market – which is coming under increasing scrutiny by competition regulators and policymakers [48, 219].

One reason for the differences in AdId access might be the restrictions set by the platforms themselves. Apple is taking steps against the use of the AdId, which is often linked to advertising. On submission to the Apple App Store, app publishers have long had to declare that their app only uses the AdId for certain, specific reasons related to advertising. Additionally, Apple allows iOS users to prevent all apps from accessing the AdId, and even asks for explicit opt-in as of iOS 14.5. Google does not currently allow all users to prevent apps from accessing the AdId on Android (although the company is rolling out a new opt-out framework).

Android Permission	Apps (%)	Opt-in?	iOS Permission	Apps (%)	Opt-in?
INTERNET	98.7	x	PhotoLibrary	58.0	✓
ACCESS_NETWORK_STATE	94.5	x	Camera	56.3	✓
WAKE_LOCK	70.0	x	LocationWhenInUse	47.7	✓
WRITE_EXTERNAL_STORAGE	63.4	✓	LocationAlways	31.4	✓
READ_EXTERNAL_STORAGE	41.4	✓	PhotoLibraryAdd	27.4	✓
ACCESS_WIFI_STATE	40.0	x	Microphone	26.2	✓
VIBRATE	35.8	x	Calendars	25.2	✓
RECEIVE_BOOT_COMPLETED	26.5	x	LocationAlwaysAndWhenInUse	16.8	✓
ACCESS_FINE_LOCATION	26.1	✓	BluetoothPeripheral	16.4	✓
ACCESS_COARSE_LOCATION	24.8	✓	Contacts	16.1	✓
READ_PHONE_STATE	21.5	✓	Motion	8.0	✓
CAMERA	21.4	✓	Location	7.5	✓
FOREGROUND_SERVICE	12.1	x	AppleMusic	7.1	✓
GET_ACCOUNTS	10.1	✓	BluetoothAlways	6.8	✓
RECORD_AUDIO	9.7	✓	FaceID	6.1	✓

(a) Most common permissions on iOS and Android.



(b) Percentage of apps requesting opt-in permissions. iOS apps consistently included more than Android.

Figure 6.3: Top permissions on Android and iOS. All permissions on iOS require opt-in, only 'dangerous' ones on Android.

6.3.2.2 Permissions

Most prevalent permissions. Figure 6.3a shows the most prevalent permissions on both platforms, and whether these require opt-in. The most common permissions on Android are `INTERNET` and `ACCESS_NETWORK_STATE`, both requested by more than 90% of apps and related to Internet access. A similar permission does not

exist on iOS. The most common 'dangerous' permissions (requiring user opt-in) on Android are related to storing and reading information on the external storage, `WRITE_EXTERNAL_STORAGE` and `READ_EXTERNAL_STORAGE`. Such external storage exists on iOS as well, but apps access it through a system-provided 'Files' dialog. `PhotoLibrary` (for photo access) is the most common permission on iOS. Although a similar permission (`CAMERA`) exists on Android, apps do not have to request it, but can rather invoke the camera application on the phone to take a picture directly. This potentially explains some of the differences in the number of camera-related permission requests between Android and iOS. The iOS `PhotoLibrary` permission was an outlier from the overall observation that iOS apps needed more permissions and was as prevalent (about 60% of apps) as the `WRITE_EXTERNAL_STORAGE` permission on Android, possibly because the most common usage of file access is processing photos (e.g. in social media or photography apps). Access to external storage can be a privacy risk since it can enable unexpected data exposure and tracking across apps [102]. Because of this, Google has been restricting access to external storage ever more with recent versions of Android and Apple has never allowed direct access to file storage.

Cross-platform permissions. All cross-platform permissions were more common on iOS than on Android, see Figure 6.3b. The most common were Camera and Location. Both were included by about 50% of iOS apps (Camera 56.3%, Location 49.2%), and less than a third of Android apps (Camera 21.2%, Location 28.0%). iOS apps also accessed the Calendar and Contacts permissions more often than Android apps (25.2% vs. 3.2% for Calendar; 16.1% vs. 6.4% for Contacts). Note that Android differentiates between read and write access for the Contacts and Calendar permission. The studied Android apps with Calendar access usually had both read (95.0%) and write (94.5%) access. Of those with Contacts access, 97.6% had read and 47.1% write access, underlining the potential value of separating read and write permissions. Motion was the least common cross-platform permission, present in 8.0% of iOS apps and 1.4% of Android apps.

6. Choice between iOS and Android

To seek further explanations on why iOS apps request camera and location access more frequently than Android apps, we first checked the categories of the apps in our dataset. Our intuition was that iOS apps would more frequently fall into categories related to photography and navigation, but that was not the case. Next, we checked the required permissions of the top 15 tracking libraries. However, we did not find any differences, except for AdColony, which requests different permissions on Android and iOS, but has a relatively small market share. Finally, we measured how many apps mentioned the terms 'photo' or 'camera' in their description on the respective app store. Including only those descriptions that were in English (according to the `langdetect` Python library [220]), 14.0% of apps on Android and 11.8% of apps on iOS mentioned either term. However, the median length of descriptions in English was substantially higher on Android (1032 characters) compared to iOS (761 characters) and only 72.7% of iOS app descriptions were actually in English (81.6% on Android), making it difficult to interpret these observations.

Summary. Android has many permissions that have no equivalent on iOS, and thus Android apps can *appear* to be more privileged than their iOS counterparts, but on closer examination, they are simply asking for permissions to access resources which are not restricted on iOS (e.g. Internet access and network state). Further, iOS apps showed substantially higher levels of cross-platform permissions that both Apple and Google deem as particularly dangerous and require user opt-in. This can be a reason for concern. Once a permission is granted, an app can usually access sensitive data anytime without the user's knowledge.

From our analysis, it does not seem like the distribution of apps on the app stores, or the different permission requirements of tracking libraries on either platform are the main drivers behind the observed differences in permission use. Instead, there are a range of architectural differences between the platforms that might lead to increased use of opt-in permissions on iOS. One important factor might be that Android allows for deeper integration between apps, through its powerful *intent system*. Android apps can call specific functionality of other apps, and listen for return values. In the past, Android apps have also been observed to

use side channels to circumvent the permission system [102], which underlines the potentially deep integration between Android apps. By contrast, iOS only allows for very limited cross-app communication. This might mean that a higher number of dangerous cross-platform permissions on iOS might actually be positive for privacy, since it reflects higher encapsulation of apps.

There has also been a wealth of research into Android's permission system in the past, and much less so on iOS; this, in conjunction with disclosures of permissions on the Google Play Store (but traditionally not so on the App Store), might have made Android developers more cautious about declaring permissions – particularly those that require explicit opt-in.

In sum, there are a variety of aspects – including differences in software architecture, developer attitudes and practices, and socioeconomics of end-users – that drive permission use on either platform. We leave it for future work to disentangle these aspects further.

6.3.3 Data Sharing

6.3.3.1 Before Consent

We now turn to data sharing in apps' network traffic, before any user interaction. While in this section we do not analyse what personal data is shared, tracker companies necessarily receive the user's IP address from any connections, which itself can classify as personal data under EU law [221] and can be used for tracking purposes [222]. Our results are shown in Figure 6.2b.

The average app on both platforms contacted similar numbers of tracking domains (2.7 on Android, and 2.4 on iOS). 18.6% of Android apps and 31.5% of iOS apps did not contact any tracking domains at the app start. The most popular domain (`googleads.g.doubleclick.net`) on Android was related to Google's advertising business – contacted by 37.6% of Android apps, and 11.9% on iOS. The most popular domain on iOS was related to Google's analytics services (`app-measurement.com`) – contacted by 30.7% of apps on iOS, and 36.4% on

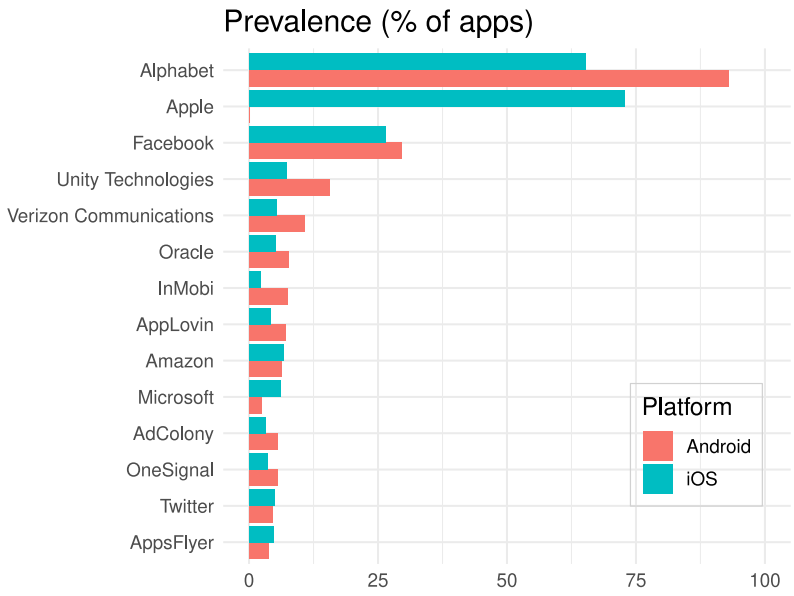
Android. Facebook services were contacted by more iOS apps (21.2%) than on Android (17.2%). Some iOS apps additionally exchange data about app installs (`itunes.apple.com`) and ad attribution (`ca.iadsdk.apple.com`) with Apple. These services are unique to the Apple ecosystem, and do not exist on Android. As observed in the previous section, advertising services seem more popular on Android than on iOS, by a factor of roughly 2 (e.g. `*.doubleclick.net`, `*.googlesyndication.com`, `unityads.unity3d.com`).

Widespread tracking without the legally required consent. Overall, we find that data sharing with tracker companies before any user interaction is common on both platforms. However, EU and UK law usually requires user consent before third-party tracking can take place (see Chapter 5). This suggests potentially widespread violations of applicable data protection law (in 81.44% of Android apps, and 68.46% of iOS apps). While most of this data sharing can be attributed to Google, other companies, such as Facebook and Unity, also receive data for tracking purposes. Moreover, tracking by Google also happens widely on iOS where, unlike on Android, a user would not have given consent as part of the device set-up process.

6.3.3.2 PII Exposure

We found that more Android apps shared the AdId over the Internet (55.4% on Android, and 31.0% on iOS). The reduced sharing of the AdId on iOS might be related to the reduced prominence of AdId access in iOS apps as found in our static analysis, and the stricter policies by Apple regarding AdId use (see Section 6.3.2.2). 85.1% of Android and 61.4% of iOS apps shared the model and phone name over the Internet, which can be used as part of device fingerprinting.

Android apps also shared other system identifiers, including the Android ID (18.2% of apps), the IMEI (1.3% of apps), the Serial number (1.1% of apps) and the Wi-Fi MAC Address (0.6% of apps). We note that these identifiers are no longer accessible as of Android 10. We did not find equivalent identifiers in iOS network traffic; iOS has long deprecated access to permanent identifiers (UDID with iOS 6 in 2012 and MAC Address with iOS 7 in 2013).



	Median	Mean	Q1	Q3	Count > 10	None
Android	1	2.3	1	2	2.77%	6.86%
iOS	2	2.6	1	3	1.82%	15.28%

Figure 6.4: Root companies that are ultimately behind tracking.

6.3.4 Tracker Companies

Owners of tracking technology. Since many tracker companies belong to a larger consortium of companies (see Figure 3.1 for the example of Verizon), we now consider what parent companies ultimately own the tracking technology, i.e. the *root companies* behind tracker companies. We report these root companies by combining the observations from our static and traffic analysis, and checking against our X-Ray 2020 (see Section 6.2.4.

Figure 6.4 shows both the prevalence of root parents (i.e. their share among all apps), as well as descriptive statistics. The median number of companies was 1 on Android, 2 on iOS. This reflects the fact that Google is prominent in data collection from apps on both platforms, but Apple only on iOS. The maximum number of companies was 21 on Android, and 23 on iOS.

A large percentage of apps share data with one or more tracker companies ultimately owned by Alphabet, the parent company of Google, as can be seen in Figure 6.4. This company can collect data from nearly 100% of Android apps,

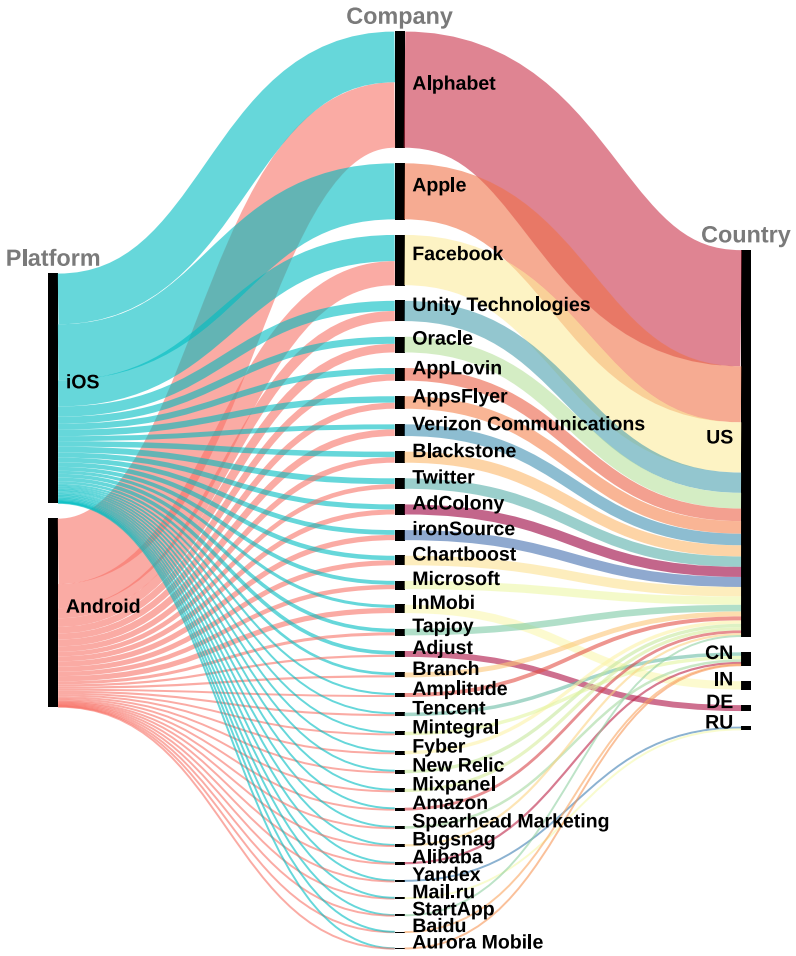


Figure 6.5: Visualisation of third-party tracking across platforms, root companies, and the jurisdictions of these root companies.

and has its tracking libraries integrated into them. Apple can collect tracking data (mainly about users' interactions with in-app ads and purchases) from more than two-thirds of apps. The next most common is Facebook, which has a similar presence on both platforms, and slightly more so on Android. Tracker companies owned by Unity and Verizon can be contacted by roughly twice as many Android apps as iOS ones. Beyond these larger companies, a range of smaller specialised tracker companies (including InMobi, AppLovin, AdColony) engage in smartphone tracking. These can potentially pose unexpected privacy risks, since they attract much less scrutiny from regulators and the interested public.

Countries of tracker companies. Based upon the X-Ray 2020 database, which contains company jurisdictions, we now can analyse in what countries the companies

behind app tracking are based (including both subsidiary and root parent). This is visualised in Figure 6.5 (root parents only). It should be noted that while some large companies (like Google and Apple) have subsidiaries in Europe, we tended to observe direct data sharing with the parent company (i.e. which are mostly US-based).

The US is the most prominent jurisdiction for tracker companies. 93.3% of Android apps and 83.5% of iOS apps can send data to a US-based company. The next most common destinations are China on iOS (9.5% of iOS apps; 4.8% on Android) and India on Android (7.45% of Android apps; 2.23% on iOS). These destinations highlight the global distribution models of both the Apple and Google ecosystem. While Google Play has a large user base in India, it is not available in China where instead numerous other app stores compete [59]. Conversely, the Apple App Store is available in China, and is the only authorised app marketplace on iOS. Germany and Russia are the only other countries whose root tracker companies reach more than 2% of apps on iOS or Android.

While we downloaded apps from the UK app store, the most commonly contacted tracking countries are based outside the UK and EU. This can give rise to potential violations of EU and UK data protection law, since the exchange of personal data beyond the EU / UK is only legal if special safeguards are put in place, or an *adequacy decision* by the European Commission (or its UK equivalent) exists [4, 223]. However, such adequacy decisions do not exist for the three most common jurisdictions of tracker companies, namely the US, China and India. According to our data, the exchange of data with companies based in countries without an adequacy decision seems similarly widespread on both Android and iOS.

6.3.5 Cross-Platform Apps

Many previous studies pursuing cross-platform app analysis (i.e. analysing both Android and iOS apps) focused on those apps that exist on both platforms. However, there has been limited discussion of how the characteristics of those *cross-platform apps* might differ from those of the average app on either platform. Our data suggests notable differences.

Platform	Android			iOS		
Category	All	Cross-Platform	Children	All	Cross-Platform	Children
Total Number	11 876	1 623	371	11 964	1 534	109
Root Tracker companies	2.3	2.8	2.7	2.6	**3.3**	2.4
Permissions (Cross-Pltf.)	11.0 (0.8)	14.3 (1.2)	6.9 (0.2)	3.7 (2.0)	4.0 (**2.1**)	2.7 (1.4)
Location Permission	28.0%	41.1%	3.8%	49.2%	**53.1%**	26.6%
AdId access (in traffic)	86.1% (55.4%)	84.4% (**64.3%**)	**89.8%** (59.3%)	42.7% (30.9%)	49.9% (38.3%)	50.5% (24.8%)

Table 6.2: Comparative statistics for all, cross-platform and children's apps on iOS and Android. Since iOS and Android have permissions of different kinds and absolute numbers, we also provide means both for all and cross-platform permissions (as defined in Section 6.2.2.3). Column-wise maxima in bold.

Table 6.2 shows a comparison between cross-platform and all apps across a range of privacy indicators (and also children's apps, which are discussed in the next Section). All privacy indicators show worse properties than for all apps: data sharing with tracker companies, the presence of permissions, potential access to location, and the communication of the AdId over the Internet were increased among cross-platform apps. On Android, cross-platform apps could share data with 2.8 companies on average, compared to 2.3 in the total Android sample. The difference of 0.5 in the average number of companies is statistically significant ($p < 0.001$, permutation test with 10,000 permutations). Their iOS counterparts could share with 3.3 companies on average (compared to 2.6 in the total iOS sample, $p < 0.001$). The mean number of permissions increased from 11.0 to 14.3 on Android ($p < 0.001$), and from 3.7 to 4.0 on iOS ($p = 0.003$). When focusing on cross-platform permissions, the figure increased from 0.8 to 1.2 on Android ($p < 0.001$), and from 2.0 to 2.1 on iOS ($p = 0.007$). Apps had a similar level of AdId access on Android than across all apps ($p = 0.06$). However, more apps (64.3% in cross-platform apps compared to 55.4%) were observed to share the AdId over the Internet ($p < 0.001$). Similarly, the proportion of apps that share the AdId over the Internet increased from 30.9% to 38.3% on iOS ($p < 0.001$).

The reason for the higher amount of tracking in cross-platform apps may be due to increased popularity, and thereby heightened financial interest in data collection for advertising and analytics purposes. This makes it not only more valuable to use user data for advertising and other purposes, but also to develop an app for both platforms in the first place. Indeed, manual analysis showed that among the

top 100 apps from the UK app stores on Android and iOS, 92% existed for both platforms. The more popular an app, the more likely it seems to be available on both platforms and the more likely it is to use a greater number of tracking services.

6.3.6 Apps for Children

Children enjoy special protections under data protection laws in many jurisdictions, including COPPA in the US and the GDPR in the EU and UK. Among other legal requirements, US, EU and UK law require parental consent for many data collection activities involving children. The UK's Age appropriate design code explicitly prohibits the use of profiling technologies without prior consent [224]. In addition to the legal requirements, Apple and Google impose contractual obligations on children's apps on their app stores. As such, the study of children's apps not only allows us to assess the practices of apps aimed at particularly vulnerable users, but also serves as a useful case study for the efficacy of privacy rules imposed by policymakers and app platforms. Both app stores offer a dedicated section for children's apps, known as the *Kids* category on the Apple App Store and the *Designed for Families* program on the Google Play Store. 109 iOS apps (0.9%) and 371 Android apps (3.1%) from our dataset fell into these categories. While this dataset is much smaller than in the previous sections, our analysis of this subset suggests that worrying privacy practices are not absent from children's apps, see Table 6.2.

Tracking. On average, tracking – in terms of the root companies present – was more widespread in Android apps for children than across all apps ($p = 0.02$, using a permutation test with $10,000$ permutations and the difference in mean as our test statistic), but not so for iOS ($p = 0.41$, 95% CI $[2.05, 2.76]$). Most of this tracking is related to analytics purposes on iOS. 84.4% of iOS apps contained Apple's SKAdNetwork (compared to 69.9% across all iOS apps), which is used for ad attribution. The next most common tracking libraries in children's apps on iOS are Google Firebase Analytics (40.4%, compared to 54.7% on Android), Google Crashlytics (22.0%, compared to 14.0% on Android), and the Facebook SDK (13.8%, compared to 17.8% on Android). As for Android, the most commonly contacted

domain (50.4% of apps) was `googleads.g.doubleclick.net`, which belongs to Google's advertising business. 71.7% of Android children's apps contained Google AdMob (compared to 14.7% on iOS); Unity3d Ads was present in 27.0% of Android children's apps (compared to 6.42% on iOS).

AdId. The increased prevalence of advertising-related tracking in children's apps on Android is consistent with the fact that more children's apps on Android were observed to share the AdId over the Internet compared to all apps (59.3% compared to 55.5%, $p = 0.14$, 95% CI $[54.3, 64.3]$), but not so on iOS (24.8% compared to 30.9%, $p = 0.17$, 95% CI $[16.4, 33.0]$) – these observed differences were not statistically significant, but the 95% confidence intervals still point to common sharing of this identifier over the Internet. The differences in AdId access between the platforms, and potentially the lower proportions of children's apps on the App Store might stem from a differing stringency of privacy rules on the two app stores. Apple started to restrict third-party data collection from children's apps [225] from June 2019 onwards. Children's apps 'may not send personally identifiable information or device information to third parties' [226], which includes personalised advertising. While the Google Play Store also bans personalised ads in children's apps, the sharing of personally identifiable or device information is not expressly prohibited [227].

Permissions. Permission use was, on average, lower than across all apps ($p < 0.01$), which could hint at improved privacy properties in children's apps. At the same time, more than one-quarter of children's apps on iOS (26.6%, compared to 49.2% across all apps, $p < 0.001$), and 3.8% (compared to 28.0% across all apps, $p < 0.001$) on Android request location access. These results reflect the fact that Google Play apps in the Family category are not allowed to access location [227].

It remains unclear from our data 1) why a minority of Android apps still declare the location permissions in their app manifest, and 2) whether some apps might obtain user location in other ways, e.g. through side-channels [102].

Conclusions. The study of children's apps revealed that many share data, including unique identifiers, with tracker companies – both on Android and iOS. The

sharing of data with advertising services, including unique user identifiers, tended to be more common on Android than on iOS ($p < 0.001$). At the same time, iOS apps contained the location permission seven times more often than their Android counterparts ($p < 0.001$), which can lead to unexpected disclosures of GPS data from children. Data sharing with third parties often takes place without the necessary parental consent, and despite privacy laws and the policies of the platforms. Not all comparisons between the subset of children's apps and all apps were statistically significant, but even where this was not the case, the reported 95% confidence intervals still underlined that worrying data practices are common in children's apps.

6.4 Limitations

It is important to highlight certain limitations of our methodology. We do not cover all apps available in each app store, only a (large) subset of free apps. Our sampling method relies on the app stores' search functionality, which might be biased differently on each platform. We excluded apps that were last updated before 2018, assuming that these are not widely used anymore. The results of our code analysis must be interpreted with care, since not all parts of an app might be invoked in practice – an inherent limitation of this type of analysis. We used off-device network analysis, which may wrongly attribute some communications; we reduced the impact of this by disabling pre-installed apps if possible. We also used jailbreaking on iOS and rooting on Android to circumvent certificate validation, which might make some apps alter their behaviour. For network analysis, we used a phone running Android 7, which was somewhat outdated at the time of data collection, but still widely used [228]. Compared to other research studies, we did not interact with the studied apps, so as to analyse data sharing without user consent. We also did not analyse interdependent privacy, i.e. how information disclosed from one individual might affect someone else. In all parts of our analysis, we consider all apps equally, regardless of popularity [47] and usage time [135], both of which can impact user privacy. Likewise, we treat all tracking domains, libraries and companies equally, though they might pose different risks to users.

6.5 Conclusions & Future Work

While it has been argued that the choice of smartphone architecture might protect user privacy, no clear winner between iOS and Android emerges from our analysis. Data sharing for tracking purposes was common on both platforms. Android apps tended to share the AdId, which can be used for tracking users across apps, more often than iOS apps ($p < 0.001$). Permissions, which both Apple and Google deem as particularly dangerous and require user opt-in, were more common among iOS apps (although Android also has a greater range of permissions deemed 'not dangerous' and do not require opt-in).

Compliance issues. Across all studied apps, our study highlights widespread potential infringements of US, EU and UK data protection and privacy laws. Apps widely use third-party tracking without user consent, lack parental consent before sharing PII with third parties in children's apps, share more data with trackers than necessary, and send personal data to countries without an adequate level of data protection.

A fundamental compliance issue is the lack of transparency around apps' data practices. Data protection law obliges apps to disclose their data practices adequately (e.g. Article 13 GDPR). Privacy policies are one way to do this, but are often inadequate [42, 58, 70, 73]. At the same time, design decisions by Apple and Google hinder the interested public from independently assessing the privacy practices of apps. Apple even applies encryption to all iOS apps and widely uses proprietary technologies, thereby driving researchers analysing iOS apps into legal grey areas. On Android, Google has banned the installation of root certificates in unmodified versions of Android (which is necessary to assess apps' network communications), enabled app obfuscation in release builds by default, and has been taking measures against those who modify their Android device with its SafetyNet (even if this is for research purposes). These new hurdles to app privacy research are in potential conflict with the transparency obligations under data protection and privacy laws.

Apple and Google's conflicts of interest. Since the platforms take a share of any sales through the app stores (up to 30%), both Apple and Google have a natural interest in creating business opportunities for app publishers, and letting them collect data about users to drive such sales. Apple's AdId policies might actively encourage certain app monetisation models to its own benefit (Section 6.3.2.1). Our study also underlined the high market share of Google in mobile display advertising, which itself relies on the collection of user data. Google Ads was potentially present in more than half of apps with ads on both iOS and Android (Section 6.3.2.1).

The study of children's apps further illustrated the conflicts of interest that app platforms face between user privacy and revenues. Both platforms have policies to limit data collection and advertising in children's apps. Despite this, access to unique device identifiers, specifically the AdId, and access to user location was still common in children's apps. 27% of children's apps on iOS could request the user location, and 4% on Android. About 59% of Android apps shared the AdId with third parties over the Internet, 25% on iOS. This can be used to build fine-grained profiles about children, putting them at risk [30].

As a result of these conflicts of interest, Google and Apple's business practices are being investigated by competition regulators worldwide, including in the US [200, 201], the EU [202], Germany [219], and the UK [48]. Indeed, the US Department of Justice is investigating potentially anti-competitive and illegal contracts between the two companies [201].

Suggestions. App platforms are well-positioned to protect user privacy [82, 97, 108], but targeted regulation of app platforms remains largely absent (see Section 2.4.4). This stresses the need for increased transparency around apps' practices [229]. More transparency could also help build trust around the changing takes by platforms on user privacy, including the scanning of users' photo libraries for Child Sexual Abuse Material (CSAM) as proposed by Apple in 2021 [230, 231]. In the meantime, transparency around the privacy practices of apps will remain a challenging target to analyse, as will creating accountability for privacy malpractices. The tools developed in this Chapter seek to foster discussion on regulatory and

transparency issues around app privacy, and we share all our tools and data publicly to support such work at https://platformcontrol.org/.

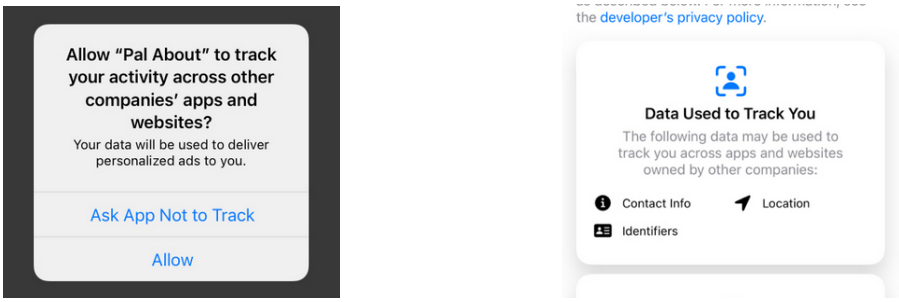
Future work. An important field for further study is the development of a cross-platform app instrumentation tool. Further research is also needed to develop a holistic approach to assess compliance within mobile apps. Since Apple introduced new wide-ranging privacy measures with iOS 14, the analysis of the impact of these changes is another important piece of follow-up work and conducted the next Chapter.

7

Impact of iOS App Tracking Transparency and Privacy Labels

Starting with iOS 14, Apple introduced two new measures to improve the privacy protections on iOS: the *App Tracking Transparency* (ATT) framework and *Privacy Nutrition Labels*. These new measures are introduced in response to the heightened privacy expectations of iOS users (see Section 2.1.3) as well as increased regulatory pressure (see Section 4.1).

Under the ATT, iOS apps must now ask users for explicit permission before tracking them (see Figure 7.1a). If an iOS user asks an app not to track, then this has the direct effect that this app cannot access the Identifier for Advertisers (IDFA) anymore. The IDFA is a random, unique identifier provided by the operating system to apps for tracking users across different apps and multiple sessions of a single app. Additionally, apps are obliged to stop certain tracking practices under Apple's App Store policies (more in Section 7.3). Preliminary data suggests that the vast majority of users (between 60% and 95%) choose to refuse tracking when asked for it under the new system [158, 159, 232]. Although users could previously opt-out from the use of the IDFA, the use of this feature was low, since it was off by default and not very visible in the system settings of iOS devices [233].



(a) App Tracking Transparency (ATT). **(b)** Privacy Nutrition Label.

Figure 7.1: Overview of Apple's new privacy measures, introduced with iOS 14 [24].

While potentially good for user privacy, the ATT has been reported to have vastly increased Apple's share of advertising on iOS – as part of its Apple Search Ads on the App Store – and to have decreased the share and efficacy of ads from competing companies. An important reason for this, as argued by Eric Seufert and others, is that Apple's own tracking technologies may not fall under Apple's definition of tracking [161]. It has also been reported that, as a result, many marketing companies have shifted advertising budgets from iOS to Android [160]. The Financial Times estimated that the loss for leading tech companies from the new policy would be around $10bn [234], but also reported that companies deemed the 'effect of Apple's privacy changes was less than feared' [235]. Apple's new rules might represent an important shift in the mobile advertising ecosystem towards more user privacy and market concentration. They may also prompt a rise in paid apps and in-app purchases [236], and thereby particularly affect those individuals who are already worse off financially.

Despite Apple's new rules, some apps may try to circumvent them so as to continue to engage in the lucrative collection and sharing of personal data. This might sometimes be in violation of Apple's App Store policies, but both the automatic and manual review on the App Store can sometimes fail to pick up such violations and not enforce its policies [33, 82, 237–239]. The refusal to respect users' consent to tracking might not only be in violation of Apple's own rules, but also defy the expectations of many end-users and potentially violate existing legal requirements relating to consent to tracking.

In addition to the changes relating to the ATT, app developers must now self-declare what types of data they collect from users, and for what purposes – called *Privacy Nutrition Labels* [240, 241], see Figure 7.1. As such, these labels aim to make it easier for end-users to understand the data practices of apps, instead of having to study lengthy privacy policies, which few users do [9]. There is, however, a risk that many users may just ignore the new (and potentially oversimplified) privacy labels (as they commonly do with privacy policies [9]), gain a false sense of security, or not understand the consequences for their privacy (which tends to be highly individual [52]), and that developers may not honestly self-declare their actual data practices [242]. Despite these concerns, the labels have the potential to shift developers' existing data practices towards being more privacy-preserving, through increased transparency and end-user awareness.

Based on the above observations, this Chapter analyses the following research questions:

1. What impact have the ATT and Privacy Nutrition Labels had – thus far – on tracking, particularly on the extent and quality of tracking?

2. To what extent do apps disclose their tracking practices in their Privacy Nutrition Labels?

3. What implications do the ATT and Privacy Nutrition Labels have for the power relations between the actors in the digital advertising system, including mobile OS providers, digital advertisers, app developers and marketers?

To analyse these questions, this Chapter analyses privacy in 1,759 iOS apps, for each of which we downloaded two versions: one from before Apple's new rules and one that has been updated since. We use a combination of app code and network analysis to gain rich insights into the data practices of the studied apps.

Contributions. By answering the above questions, this Chapter provides important evidence around the marketing claims of Apple regarding privacy. We further provide the first real-world evidence of apps using fingerprinting to agree

on a mutual user identifier, thereby sidestepping Apple's new privacy rules. This Chapter also underlines the conflicts of interest that app platforms like Apple face in regulating their app ecosystems, and contributes to our understanding of how tracking might develop in the future.

Structure. The remainder of this Chapter is structured as follows. We first introduce our app selection and analysis methodology in Section 7.1. We turn to the results from our app code and network analysis in Section 7.2. We discuss our findings in Section 7.3 and the limitations of our study in Section 7.4. We conclude the Chapter and outline direction for future work in Section 7.5.

7.1 Methodology

In this Section, we describe our analysis methodology, which follows the one previously used for the comparative analysis of iOS and Android apps' privacy practices in the previous Chapter 6.

7.1.1 App Selection and Download

For the selection of apps, we revisited the same 12,000 iOS apps as in the previous Chapter 6. We re-downloaded those apps that were updated to comply with Apple's ATT and privacy label rules, in October 2021. This resulted in a dataset of 1,759 *pairs* of apps, one from before iOS 14 and one from after. This number of apps is comparatively small because many apps had not yet been updated since the new rules, while some other apps had been removed from the store (2,713 out of 12,000 apps were not available on the App Store anymore). We additionally downloaded the Privacy Nutrition Labels for the newly downloaded apps.

7.1.2 App Analysis

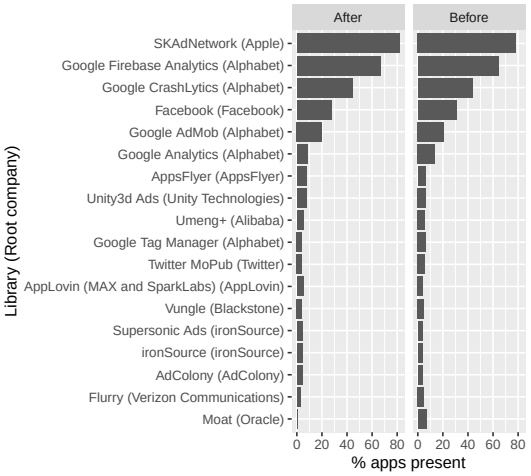
For our further analysis of apps, we executed every app on a real device – one iPhone SE 1st Gen with iOS 14.2, and one with iOS 14.8 – for 30 seconds without user interaction. We captured apps' network traffic using the tool `mitmdump`. We

disabled certificate validation using `SSL Kill Switch 2`, after gaining system-level access on both iPhones (known as 'jailbreak'). On the iPhone with iOS 14.2, we did not opt-out from ad personalisation from the system settings, thereby assuming user opt-in to use the IDFA (reflecting the assumption that many users, who would reject tracking, do not do so because the option is in the less prominent settings on the OS [36]). On the iPhone with iOS 14.8, we asked all apps not to track from the system settings. To identify the presence of tracking libraries, we extracted the names of all classes loaded by each app using the tool `Frida` [243] and checked them against a list of known tracker class names, curated in the previous Chapter 6. As in the previous Chapter 6, we also analysed the privacy settings provided by some of the most prominent tracking libraries: Google AdMob, Facebook, and Google Firebase. Beyond analysing tracking in apps, we again obtained a list of permissions that apps can request. These permissions are different to the new privacy labels, which do not affect the runtime behaviour of apps. We extracted apps' permissions by automatically inspecting the manifest file that every iOS app must provide.

7.2 Results

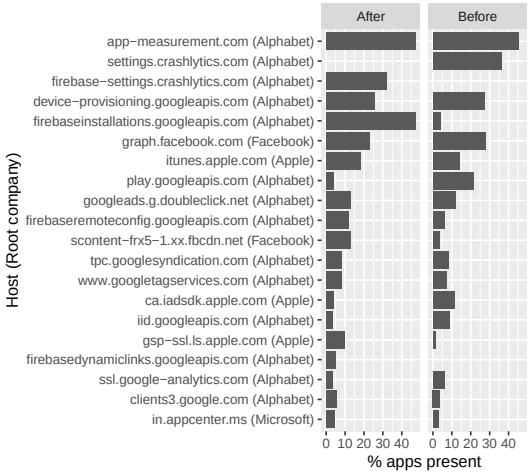
In this Section, we present our findings from analysing two versions – one from before and one from after the release of iOS 14 and the ATT – of 1,759 iOS apps. We analysed 199.6 GB of downloaded apps, extracted 3.2 GB of information about classes in apps' code, and collected 3.9 GB of data in apps' network traffic. Installing and instrumentation failed for 74 iOS apps; we excluded these apps from our subsequent analysis and focused on the remaining 1,685 apps.

First, we focus on the tracking libraries found in the code analysis (Section 7.2.1) and whether or not they were configured for data minimisation (Section 7.2.1.1). Following up, in Section 7.2.2, we analyse apps' access to the IDFA (which is now protected by the ATT) and also their permissions. Next, in Section 7.2.3, we report on the data sharing of apps before consent is provided, with a particular focus on whether apps that are instructed not to track actually do so in practice.



	Median	Mean	Q1	Q3	Count > 10	None
Before	3	3.7	2	5	4.75%	13.61%
After	3	3.6	2	4	4.75%	12.48%

(a) Top tracking libraries in app code.



	Median	Mean	Q1	Q3	Count > 10	None
Before	3	4.0	1	6	4.75%	13.43%
After	4	4.7	2	7	7.19%	10.70%

(b) Top tracking hosts contacted at the first app start.

Figure 7.2: Third-party libraries (integrated in apps, but not necessarily activated) and contacted tracking domains of apps, as well as the companies owning them (in brackets). Shown are the top 15 tracking libraries and domains for before and after the new privacy changes under iOS 14.

Lastly, in Section 7.2.4, we check whether and to what extent apps disclose their tracking practices in their Privacy Nutrition Labels.

7.2.1 Tracking Libraries

Apps from both before the ATT and after widely used tracking libraries (see Figure 7.2a). The median number of tracking libraries included in an app was 3 in both datasets. The mean before was 3.7, the mean after was 3.6. 4.75% of apps from before ATT contained more than 10 tracking libraries, compared to 4.75% after. 86.39% contained at least one before ATT, and 87.52% after.

The most prominent libraries have not changed since the introduction of ATT. The top one was the SKAdNetwork library (in 78.4% of apps before, and 81.8% after). While part of Apple's privacy-preserving advertising attribution system, this library discloses information about what ads a user clicked on to Apple, from which Apple could (theoretically) build user profiles for its own advertising system.

Following up with Apple about this potential issue (by exercising the GDPR's *right to be informed* under Article 13), they did not deny the fact that this data might be used for advertising, but assured us that any targeted ads would only be served to segments of users (of at least 5,000 individuals with similar interests). Google Firebase Analytics ranked second (64.3% of apps from before ATT, and 67.0% after), and Google Crashlytics third (43.6% before, 44.4% after).

Overall, Apple's privacy measures seem not to have affected the integration of tracker libraries into *existing* apps.

7.2.1.1 Configuration for Data Minimisation

Among the apps that used Google AdMob, 2.9% of apps from before and 4.5% from after chose to delay data collection. Choosing to delay data collection can be helpful for app developers, to seek consent before enabling tracking and to fulfil legal obligations. Among the apps using the Facebook SDK, there was an increase in those which delayed the sending of app events (6.7% before, and 12.5% after), an increase in those which delayed the SDK initialisation (1.0% before ATT, 2.2% after), and an increase in those which disabled the collection of the IDFA (5.0% before, 8.6% after). Among apps using Google Firebase, 0.6% permanently deactivated analytics before ATT and 0.8% after, 0.0% disabled the collection of the IDFA before and 0.6% after, and 0.6% delayed the Firebase data collection before ATT and 1.0% after.

Overall, we found that only a small fraction of apps made use of data-minimising SDK settings in their manifest files. One reason for this observation might be that some developers are not aware of these settings because tracking companies tend to have an interest in less privacy-preserving defaults regarding data collection [6, 31]. This fraction has subtly increased since the introduction of the ATT.

7.2.2 Data Access and Permissions

Most prevalent permissions. Figure 7.3 shows the most prevalent permissions before and after the introduction of the ATT. On average, there was an increase in permission use (4.3 permissions before, 4.7 after – excluding the new

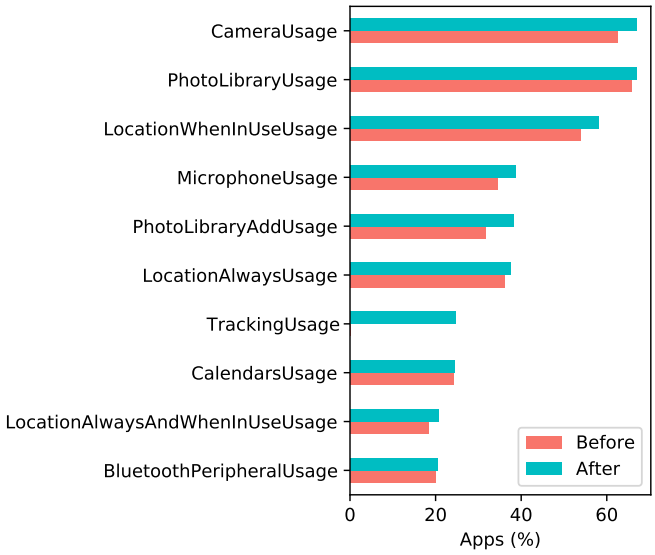


Figure 7.3: Top 10 permissions that apps can request.

`TrackingUsage` permission). Among the re-downloaded apps, `CameraUsage` (for camera access) was the most common permission (62.6% before ATT, 66.9% after), closely followed by `PhotoLibraryUsage` (65.8% before ATT, 66.9% after), and `LocationWhenInUseUsage` (53.8% before ATT, 58.0% after).

Tracking permission and access to IDFA. As part of ATT, apps that want to access the IDFA or conduct tracking must declare the `TrackingUsage` permission in their manifest. 24.7% of apps from our dataset chose to declare this permission, and might ask users for tracking. At the same time, the share of apps that contain the `AdSupport` library, necessary to access the IDFA in the app code, stayed unchanged at 50.8% of apps. This means that 50.8% of apps from after the ATT could access the IDFA on earlier versions of iOS than 14.5, but only 24.7% can on iOS 14.5 or higher.

Tracking permission and integration of tracking SDKs. The share of apps that both contained a tracking library and could request tracking varied somewhat between the used tracking library. 69.3% of the 350 apps that integrated Google AdMob declared the `TrackingUsage` permission; 78.7% of the 110 apps that integrated Unity3d Ads; 50.0% of the 116 apps that integrated Moat; and 77.3% of the 54 apps that integrated Inmobi. Whether the app is from before or

after the ATT, the vast majority of apps (between 97 and 100%) that integrated any of these tracking libraries also integrated the `AdSupport` library, and could therefore access the IDFA if running on iOS versions before 14.5.

7.2.3 Data Sharing

7.2.3.1 Before Consent

This Section analyses how many tracking domains were contacted by apps before any user interaction has taken place; the next Subsection then analyses what data was shared with trackers. Since tracking libraries usually start sending data right at the first app start [6, 36, 70, 178], this approach provides additional evidence as to the nature of tracking in apps – and without consent. Our results are shown in Figure 7.2b.

The average number of tracking domains contacted was somewhat higher for apps from after the introduction of the ATT (4.0 before, 4.7 after). The most popular domains were related to Google's analytics services: `firebaseinstallations.googleapis.com` (4.1% of apps before the ATT, 47.4% after) and `app-measurement.com` (45.2% before, 47.2% after). Since both endpoints are related to Google Firebase, the large increase in prevalence likely reflects internal restructuring of Firebase following Google's acquisitions of other advertising and analytics companies. For example, Google acquired the crash reporting software Crashlytics from Twitter in January 2017, which is reflected in our data: Google deprecated the old API endpoints (from `settings.crashlytics.com` to `firebase-settings.crashlytics.com`) from November 2020. This had the direct effect that all Crashlytics users must now also use Google Firebase. The domain `settings.crashlytics.com` was contacted by 36.4% of apps from before the ATT, and `firebase-settings.crashlytics.com` by 32.3% after the ATT. While this might point to a small difference in the adoption of Google Crashlytics, the exact same number of apps (734, 43.6%) integrated the Crashlytics library into their code, before and after the ATT. Similarly, the exact same number of apps integrate the Facebook SDK (523, 31.1%); the share of apps that contacted the associated API endpoint `graph.facebook.com` at the first start fell from 27.7%

```
{
    "sdk_version": "1.2.0",
    "bundle_id": "[Redacted]",
    "hw_model": "N69uAP",
    "kid": "[Redacted]",
    "total_storage": "30745123781",
    "country": "GB",
    "zdata": "[Redacted]",
    "app_version": "[Redacted]",
    "app_name": "[Redacted]",
    "sdk_type": "IOS",
    "storage": "14078912372",
    "zdata_ver": "1.1.0",
    "source_id": "umeng",
    "idfv": "7EBDAFC8-97BB-4FDB-B4D3-E2F4EA040B8C",
    "timezone": "1",
    "os_version": "14.8",
    "model": "iPhone8,4",
    "hostname": "MyPhone",
    "appkey": "[Redacted]",
    "idfa": "00000000-0000-0000-0000-
        000000000000"
}
```


(a) Request: Sending a range of device information to Umeng at https://aaid.umeng.com/api/postZdata.


```
{
    "aaid": "BAEC362C-49FC-494B-B0A7-175
        D990B059D",
    ...
}
```


(b) Response: Umeng returns an identifier that is shared by multiple apps, and can be used for cross-app tracking.

Information	Example	Before	After
iPhone Name	MyPhone	2.5%	4.2%
iPhone Model	iPhone8,4|iPhone SE	60.2%	74.5%
Carrier	Three	20.2%	20.2%
Locale	en_GB|en-gb	85.7%	90.1%
CPU Architecture	ARM64|16777228	13.7%	16.1%
Board Config	N69uAP	3.1%	4.5%
OS Version	14.8|18H17	79.9%	86.9%
Timezone	Europe/London	3.9%	3.4%

(c) Proportion of *all* apps that shared device information. This information can potentially be used for fingerprinting or cohort tracking.

Figure 7.4: Fingerprinting and cohort tracking in apps, even after the ATT. The fingerprinting is likely in violation of Apple's new policies and the expectations of many end-users (personal data redacted).

to 23.1%. The Google Admob SDK, too, was integrated in the same number of apps (350, 20.8%), and did not see a decline in apps that contact the associated API endpoint `googleads.g.doubleclick.net` (12.1% before, 12.9% after).

Overall, data sharing with tracker companies before any user interaction remains common, even after the introduction of the ATT. This is in potential violation of applicable data protection and privacy law in the EU and UK, which usually require prior consent (see Chapter 5).

7.2.3.2 Exposure of Personal Data

We found that 26.0% of apps from before the ATT shared the IDFA over the Internet, but none from after the ATT. In this sense, the ATT effectively prevents apps from accessing the IDFA. Despite Apple's promises, closer inspection of the network traffic showed that both Apple and other third parties are still able to engage in user tracking.

We found that iPhones continued to share a range of information with third-

parties, that can potentially be used for device fingerprinting or cohort tracking (see Table 7.4c). Only *timezone* saw a subtle decrease in the number of apps that share this information. It is not clear why apps Needed to access or shared some of this information, e.g. the carrier name (shared by 20.2% of apps) or the iPhone name (shared by 3–4% of apps). Meanwhile, some types of information, particularly the iPhone name, might allow the identification of individuals, especially when combined with other information.

In our analysis, we found 9 apps that were able to generate a mutual user identifier that can be used for cross-app tracking, through the use of server-side code. These 9 apps used an 'AAID' (potentially leaning on the term Android Advertising Identifier) implemented and generated by Umeng, a subsidiary of the Chinese tech company Alibaba. The flow to obtain an AAID is visualised in Figures 7.4a and 7.4b. As expected, the IDFA is only zeros because we used the opt-out provided by iOS 14.8; we observed, however, that the IDFV (ID for Vendors), a non-resettable, app-specific identifier was shared over the Internet, see Figure 7.4a. The sharing of device information for purposes of fingerprinting would be in violation of the Apple's policies, which do not allow developers to 'derive data from a device for the purpose of uniquely identifying it' [24]. Other experts and researchers have also voiced concerns that tracking might continue [222, 237–239].

We reported our observations to Apple on 17 November 2021, who promised to investigate the problem. We conducted a follow-up investigation on 1 February 2022, and re-downloaded and analysed a range of iOS apps. Some of the apps still continued to retrieve a unique identifier from the URL `https://aaid.umeng.com/api/postZdata`. Other apps now contact the URL `https://utoken.umeng.com/api/postZdata/v2`, and apply additional encryption (rather than just HTTPS) to requests and responses. This encrypted data had roughly the same size as before (~750 bytes for the request, ~350 bytes for the response) and the same mimetype (for the request, for the response). The issue seems thus to be present still, but has now been hidden away from the public through the use of encryption. We have tried to reproduce these experiments for a few apps on iOS 15 and higher,

Domain	Company	Apps	User ID	Locale	Model	OS Version
firebaseinstallations.googleapis.com	Google	47.4%	✓	✓		
app-measurement.com	Google	47.2%	✓	✓		
firebase-settings.crashlytics.com	Google	32.3%	✓	✓	✓	✓
device-provisioning.googleapis.com	Google	25.8%	✓	✓	✓	✓
graph.facebook.com	Facebook	23.1%	✓	✓	✓	✓
itunes.apple.com	Apple	18.3%	✓	✓	✓	✓
fbcdn.net	Facebook	13.0%		✓		
googleads.g.doubleclick.net	Google	12.9%	✓	✓	✓	✓
firebaseremoteconfig.googleapis.com	Google	11.8%	✓	✓		
gsp-ssl.ls.apple.com	Apple	9.9%	✓	✓	✓	✓
tpc.googlesyndication.com	Google	8.3%		✓		✓
www.googletagservices.com	Google	8.1%		✓		✓
clients3.google.com	Google	5.3%		✓		
firebasedynamiclinks.googleapis.com	Google	5.2%	✓	✓		✓
in.appcenter.ms	Microsoft	4.3%	✓	✓	✓	✓
play.googleapis.com	Google	4.2%	✓	✓	✓	✓
skadsdk.appsflyer.com	AppsFlyer	4.0%	✓	✓		
gsp64-ssl.ls.apple.com	Apple	3.9%		✓	✓	✓
api.onesignal.com	OneSignal	3.7%		✓		
ca.iadsdk.apple.com	Apple	3.7%	✓	✓	✓	✓

Table 7.1: 20 most common tracking domains after ATT: sharing of user identifiers with third-parties, alongside device information. Empty cells mean that we did not observe the sharing of a certain type of information, although this might still take place.

but did not observe the same behaviour; there did not exist a public jailbreak for these iOS versions at the time of our study, and similar investigations as ours were not (yet) possible on these iOS versions. There is a possibility that the issue has been fixed on iOS 15 or higher, or that we did not pick up the same behaviour in our small-scale testing (about 10 apps instead of more than 1000). However, Apple did not provide further details to us.

Analysing the top 20 most commonly contacted domains, we could confirm that installation-specific identifiers (see column 'User ID') are commonly collected alongside further device-specific information, see Table 7.1. While these installation-specific identifiers are usually randomly generated at the first app start, large tracking companies can likely still use these identifiers to build profiles of an app user's journey across apps, using their server-side code to link different identifiers together (e.g. through the user's IP address, other device information and first-party data). Companies also receive information about a user's locale (i.e. the display language), the device model, and the OS version. Such information can be used to

```
<plist version="1.0">
<dict>
    ...
    <key>dsid</key>
    <string>[Apple ID]</string>
    <key>guid</key>
    <string>[UDID]</string>
    <key>serialNumber</key>
    <string>[serial number]</string>
    ...
</dict>
</plist>
```

(a) Request of Apple App Store to https://buyitunes.apple.com/WebObjects/MZFinance.woa/wa/renewVppReceipt?guid=[UDID].

```
{
    "attributionMetadataExistsOnDevice": false,
    "toroId": "[Redacted]",
    "purchaseTimestamp": "2021-11-01T15:15:05Z",
    "adamId": 477718890,
    "attributionDownloadType": 0,
    "developmentApp": false,
    "anonymousDemandId": "[Redacted]",
    "bundleId": "ru.kinopoisk",
    "attributionKey": "[Redacted]"
}
```

(b) Request (shortended) of Apple's advertising framework to https://ca.iadsdk.apple.com/adserver/attribution/v2.

Figure 7.5: Sharing of unique user identifiers with Apple (personal data redacted).

distinguish different users connecting from the same IP address (e.g. households sharing the same Wi-Fi router) – and even across different IP addresses through the use of additional, first-party data that large tracking companies hold.

Table 7.1 does not include all the different kinds of information that we observed being sent to tracking domains because the kinds of information varied between companies. For example, Google assigned an `android_id` to an iOS app upon first contact with the company that was then used for all subsequent communication with Google's API endpoints. This identifier differed between apps, and did not seem to be used for cross-app tracking on-device (it might be on Google's servers). When contacting the domain `googleads.g.doubleclick.net`, Google collected the current system volume and the status of the silencing button. As already described above, `ca.iadsdk.apple.com` collected a `purchaseTimestamp`, that can be used to identify the user, and is not accessible by other app developers. The domain `gsp64-ssl.ls.apple.com`, belonging to Apple's location services, even collected the IP address and port that we used for proxying the network traffic through `mitmdump` as part of our analysis. We did not observe any other domains that had access to this information, underlining Apple's privileged data access. Crucially, for many of the observed transmissions between apps and servers, we could not even determine what data was sent, due to the use of encryption [238] and closed-source communication protocols.

System-level tracking by Apple. We found that iPhones exchanged a range of unique user identifiers directly with Apple, see Figure 7.5. We observed that network requests, which included various unique user identifiers and other personal data, were issued following the interaction with apps and connected to Apple's App Store and advertising technologies. While this does not allow user-level apps to gain access to these user identifiers, Apple itself can use these identifiers to enrich its own advertising services. Indeed, Apple claims in its privacy policy that it may use users' interactions with its advertising platform and with the App Store to group users into segments (of at least 5,000 individuals), and show adverts to these groups [61]. Specifically, we found that the App Store collected the UDID, the serial number of the device, the DSID (an identifier linked to a user's Apple account), and a `purchaseTimestamp`. All of these identifiers can be used by Apple to single out individual users. Crucially, the UDID has been inaccessible to app developers other than Apple since 2013 [244], but Apple continues to have access to this identifier. Moreover, Apple collects the serial number, which cannot be changed and is linked to a user's iPhone. This might be unexpected for some users. These findings are in line with previous reports that both Google and Apple collect detailed information about their users as part of regular device usage [65].

7.2.4 Disclosure of Tracking in Privacy Nutrition Labels

We now consider whether and to what extent apps (from after the introduction of iOS 14) disclose their tracking activities in their Privacy Nutrition Labels.

Among the studied apps, 22.2% claimed that they would not collect any data from the user. This was often not true: as shown in Figure 7.6, 80.2% of these apps actually contained at least one tracker library (compared to 93.1% for apps that did disclose some data sharing), and 68.6% sent data to at least one known tracking domain right at the first app start (compared to 91.4%). On average, apps that claimed not to collect data contained 1.8 tracking libraries (compared to 4.3), and contacted 2.5 tracking companies (compared to 4.2).

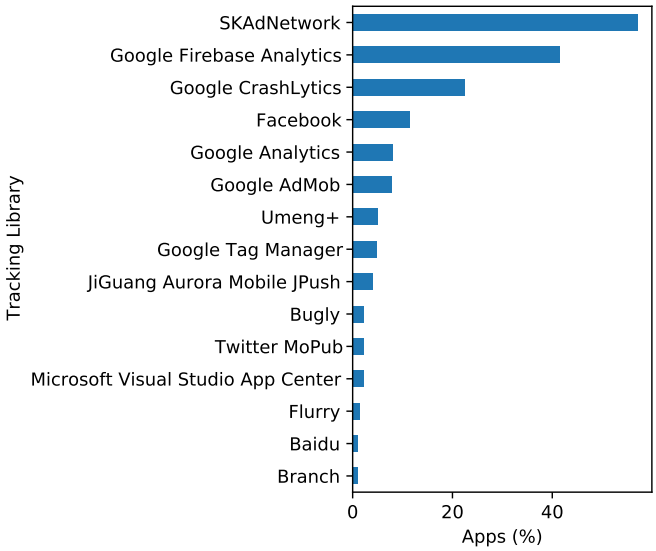


Figure 7.6: Top tracking libraries in apps that claim in their Privacy Nutrition Labels not to collect any data.

Among the 22.2% of apps claiming not to collect data, only 3 were in the App Store charts. As noted above (see Table 7.1), tracking libraries usually create a unique user identifier. Among the apps that used the SKAdNetwork, 42.0% disclosed their access to a 'User ID', 42.2% of apps using Google Firebase Analytics, 48.2% of apps using Google Crashlytics, and 53.2% of apps using the Facebook SDK. 63.2% of apps using Google Firebase Analytics disclosed that they collected any data about 'Product Interaction' or 'Other Usage Data', and about 70% of apps using the Facebook SDK, Google Analytics, or Google Tag Manager. Additionally, apps can disclose their use of 'Advertising Data': 27.5% of apps with the SKAdNetwork did so, 66.0% of apps with Google AdMob, 80.9% of apps with Unity3d Ads, and 45.4% apps with AppsFlyer.

All of this points to notable discrepancies between apps' disclosed and actual data practices. App developers might be able to address this, but they are often not fully aware of all the data that is collected through third-party tracking software [30, 31]. Conversely, Apple itself might be able to reduce this discrepancy through increased use of automated code analysis, in particular applied to third-party tracking software.

7.3 Discussion

Tracking is still widespread and reinforces the power of gatekeepers and the opacity of the mobile data ecosystem. Our findings suggest that tracking companies, especially larger ones with access to large troves of first-party data, can still track users behind the scenes. They can do this through a range of methods, including using IP addresses to link installation-specific IDs across apps and through the sign-in functionality provided by individual apps (e.g. Google or Facebook sign-in, or email address). Especially in combination with further user and device characteristics, which our data confirmed are still widely collected by tracking companies, it would be possible to analyse user behaviour across apps and websites (i.e. fingerprinting and cohort tracking). A direct result of the ATT could therefore be that existing power imbalances in the digital tracking ecosystem get reinforced.

We even found a real-world example of Umeng, a subsidiary of the Chinese tech company Alibaba, using their server-side code to provide apps with a fingerprinting-derived cross-app identifier, see Figure 7.4. The use of fingerprinting is in violation of Apple's policies [24], and raises questions about the extent to which Apple can enforce its policies against server-side code. ATT might ultimately encourage a shift of tracking technologies behind the scenes, so that they are outside of Apple's reach. In other words, Apple's new rules might lead to even less transparency around tracking than we currently have, including for academic researchers.

Privacy Nutrition Labels can be inaccurate and misleading, and have thus far not changed data practices. Our results suggest that there is a discrepancy between apps' disclosed (in their Privacy Nutrition Labels) and actual data practices. We observed that many (mostly less popular) apps gave incomplete information or falsely declared not to collect any data at all. These observations are not necessarily to blame on app developers, who often have no idea of how third-party libraries handle users' personal data [6, 30, 31]. As reported in Section 7.2.1.1, the proportion of app developers that make use of data-minimising settings of popular tracker libraries has roughly doubled, but these developers still remain a

small minority. The Privacy Nutrition Labels have not (yet) had an impact on developers' actual practices at large, but might do so in the long run by both increasing app users' privacy expectations and making app developers rethink their privacy practices [240, 241]. As they stand, the labels can be misleading and create a false sense of security for consumers.

Are the most invasive and opaque trackers tamed now? The reduced access to permanent user identifiers through ATT could substantially improve app privacy. While in the short run, some companies might try to replace the IDFA with statistical identifiers, the reduced access to non-probabilistic cross-app identifiers might make it very hard for data brokers and other smaller tracker companies to compete. Techniques like fingerprinting and cohort tracking may end up not being competitive enough compared to more privacy-preserving, on-device solutions. We are already seeing a shift in the advertising industry towards the adoption of such solutions, driven by decisions of platform gatekeepers (e.g. Google's FloC / Topics API and Android Privacy Sandbox, Apple's ATT and Privacy Nutrition Labels) (see Section 4.6 and also [245]), though more discussion is needed around the effectiveness of these privacy-protecting technologies. The net result, however, of this shift towards more privacy-preserving methods is likely going to be more concentration with the existing platform gatekeepers, as the early reports on the tripled marketing share of Apple [235], the planned overhaul of advertising technologies by Facebook/Meta and others [245], and the shifting spending patterns of advertisers [160] suggest. Advertising to iOS users – being some of the wealthiest individuals – will be an opportunity that many advertisers cannot miss out on, and so they will rely on the advertising technologies of the larger tech companies to continue targeting the right audiences with their ads.

Failure of GDPR enforcement, and power of platforms. Apple's new rules should not have a dramatic effect on the tracking of users in the EU and UK, given that existing data protection laws in these jurisdictions already ban most forms of third-party tracking without user consent (see Chapter 5). While there was a vocal outcry over Apple's new privacy measures by advertisers, the

adtech industry was aware of tightened EU and UK data protection rules since April 2016, and had plenty of time to work out a way to ensure compliance with basic provisions of the GDPR, until May 2018, including the need to seek consent from users before engaging in tracking. Broad empirical evidence, from this dissertation and other pieces of research [6, 36, 58, 70, 73, 151], suggests that apps' compliance with the GDPR is somewhat limited.

At the same time, it is worrying that a few changes by a private company (Apple) seem to have changed data protection in apps more than many years of high-level discussion and efforts by regulators, policymakers and others. This highlights the relative power of these gatekeeper companies, and the failure of regulators thus far to enforce the GDPR adequately. An effective approach to increase compliance with data protection law and privacy protections in practice might be more targeted regulation of the gatekeepers of the app ecosystem; so far, there exists limited targeted regulation in the US, UK and EU (see Section 2.4.4).

Apple's Double Standards I: Making and Enforcing App Store Policies. Our analysis shows that Apple has a competitive advantage within the iOS ecosystem in various ways. First, it both makes the rules for the App Store and interprets them in practice. This is particularly reflected in Apple's definition of tracking, which ostensibly exempts its own advertising technology [61]: 'Tracking refers to the act of linking user or device *data collected from your app* with user or device data collected from *other companies' apps, websites, or offline properties* for *targeted advertising or advertising measurement purposes*. Tracking also refers to sharing user or device data with *data brokers*.' (emphasis added) [24] In other words, for tracking to fall under Apple's definition, it must fulfil three conditions, or be done by a data broker.

Apple's definition hinges on a distinction between first-party and third-party data collection, when this is not usually the root of privacy problems. This is why the W3C defines tracking as 'the collection of data regarding a particular user's activity across multiple distinct contexts and the retention, use, or sharing of data derived from that activity outside the context in which it occurred.' [246]. Rather than *companies*, this definition is centred around different *contexts*, as is commonly

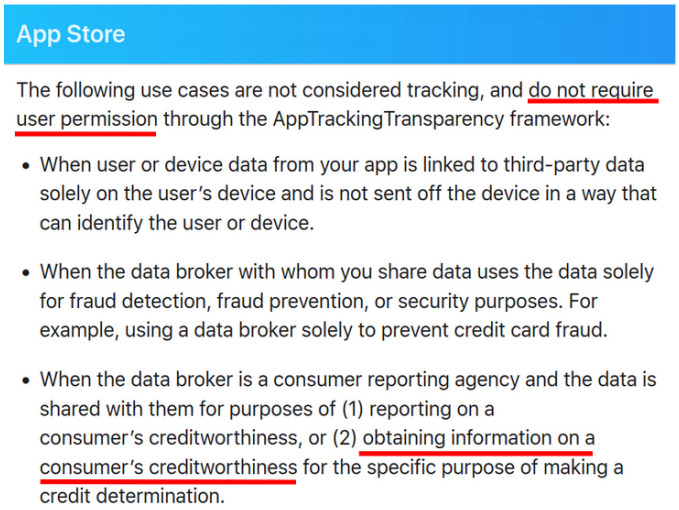


Figure 7.7: Apple exempts a list of data practices, including credit scoring, from requiring user opt-in under ATT [24] (emphasis added).

sought to be protected in privacy theory (e.g. contextual integrity [52]) and in privacy and data protection law (e.g. purpose limitation under Article 5 of the GDPR). Apple's definition of tracking might both betray the expectation of consumers who expect that tracking would stop (when first-party tracking, notably by Apple itself, continues to be allowed), and motivate other companies to consolidate and join forces leading to increased market concentration.

Apple additionally foresees a list of exempt practices [24], see Figure 7.7. These include 'fraud detection, fraud prevention, or security purposes', which might be interpreted extremely broadly by tracking companies. The exempt practices further allow tracking by a 'consumer reporting agency'. The term 'consumer reporting agency' is defined in the US Fair Credit Reporting Act (FCRA), regulating the relationship between these agencies and other 'furnishers of information' relating to consumers. By explicitly exempting credit scoring, Apple might try to avoid liability, and it might not have much choice under current US law. The exemption of credit scoring is nonetheless problematic because the use of personal data for credit scoring can have disproportionate impacts on individuals, and might be protected by other data protection and privacy laws. This might create the (false) impression for some app developers that other legal conditions do not apply, and a *false sense of security* for many consumers.

Apple's Double Standards II: Access to Data. Being the maker of the iOS ecosystem, Apple has a certain competitive advantage, by being able to collect device and user data, including hardware identifiers, that other app developers do not have access to, and use this for its own business purposes. For example, by collecting the device's serial number regularly, Apple can accurately tie the point-of-sale of its devices to activities on the device itself, and track the device lifecycle in great detail. Some of Apple's own apps, including the App Store itself, have access to this information because they are not distributed via the App Store and hence do not fall under the rules governing the App Store, including those that relate to tracking of users. These observations support the known concerns around fair competition in the App Store. Over a lack of consent to tracking by the Apple App Store in iOS 14.6, the CNIL, the French data protection authority, fined Apple 8m Euros in December 2022 [247]. Interestingly, the CNIL based its decision on violations of the ePrivacy Directive, over which it has competency (despite the one-stop shop principle of the GDPR).

7.4 Limitations

A few limitations of our study are worth noting. First, for practical reasons, we were not able to analyse all the apps in the App Store, only a reasonably large subset of free apps in the App Store's UK region. Furthermore, for the purposes of examining the effect of ATT, we only focused on apps that already existed on the App Store before iOS 14 – newly released apps may adopt different strategies. Regarding our analysis methods, our instruments are also potentially limited in several ways. The results of our static analysis must be interpreted with care, since not all code shipped in an app will necessarily be invoked in practice. We may have overestimated tracking in certain contexts, e.g., if tracking code was included but not used. In our network analysis, we performed this off-device, meaning that all device traffic was analysed in aggregate. The risk here is that we may wrongly attribute some communications to an app that was generated by some other app or subsystem on the device. To minimise this risk, we uninstalled all

pre-installed apps, and ensured no apps were running in the background. We also used jailbreaking (i.e. gained full system access by exploiting a vulnerability in the iOS operating system) to circumvent certificate validation, which might make some apps alter their behaviour. In all parts of our analysis, we consider all apps equally, regardless of popularity [47] and usage time [135], both of which can impact user privacy. Likewise, we treat all tracking domains, libraries and companies equally, though they might pose different risks to users.

7.5 Conclusions & Future Work

Overall, we find that Apple's new policies largely live up to its promises of making tracking more difficult. Tracking libraries cannot access the IDFA anymore, and this directly impacts the business of data brokers. These data brokers can pose significant risks to individuals, since they try to amass data about individuals from a wide range of contexts and sell this information to third-parties. At the same time, apps still widely use tracking technology of large companies, and send a range of user and device characteristics over the Internet for the purposes of cohort tracking and user fingerprinting. We found real-world evidence of apps computing a mutual fingerprinting-derived identifier through the use of server-side code (see Figure 7.4) – a violation of Apple's new policies [24], highlighting limits of Apple's enforcement power as a privately-owned data protection regulator [82, 108]. Indeed, Apple itself engages in some forms of user tracking (see Figure 7.5) and exempts invasive data practices like first-party tracking and credit scoring from its definition of tracking. The company was recently fined over these practices [247]. Lastly, we found the Privacy Nutrition Labels to be sometimes incomplete and inaccurate, especially in less popular apps.

Apple's privacy changes have led to positive improvements for user privacy. However, we also found various aspects that conflict with Apple's marketing claims and might go against users' reasonable privacy expectations, e.g. that the new opt-in tracking prompts would stop all tracking, that the new Privacy Nutrition Labels

would always be correct and be verified by Apple, or that Apple would be subject to the same restrictions to data access and privacy rules as other companies. There is a risk that individuals will develop even more resignation over the use of their data online if they are provided with misleading or ineffective privacy solutions [8, 18]. This resignation could in the long run undermine privacy efforts and adversely affect fundamental rights, such as the rights to data protection and privacy.

Despite positive developments over the recent months and years, especially through initiatives by Apple, there is still some way to go for app privacy. Violations of various aspects of data protection and privacy laws remain widespread in apps [6, 36, 58, 70, 73, 151], while enforcement of existing data protection laws against such practices stays sporadic. Apple's privacy efforts are hampered by its closed-source philosophy on iOS and the opacity around the enforcement of its App Store review policies. To strengthen iOS privacy, Apple has already started to prevent IP-based tracking by routing traffic to trackers via its own servers when using the iOS browser ('Privacy Relay'). As a direct response to our findings, Apple could consider extending the Privacy Relay to tracking within apps, thereby making the tracking of users through their IP address more difficult [237]. However, this would also further extend Apple's reach over the iOS ecosystem and potentially allow the company to track users even more accurately.

More generally, the key decision makers in privacy technologies must establish robust transparency and accountability measures that allow for independent assessment of any privacy guarantees and promises. This is especially true, given the current relative lack of targeted regulations for app platforms like Google Play and the Apple App Store (see Section 2.4.4). In the case of Apple, improved transparency measures must necessarily involve the phasing out of encryption of free iOS apps by default, which currently forces independent privacy researchers into legal grey areas and severely hampers such research efforts (see Section 6.1.1). This is why most previous privacy research focused on Android and the last large-scale privacy study into iOS apps had been conducted in 2013 [33], until the recent release of the method used in this dissertation (see previous Chapter 6).

We conclude that the new changes by Apple have traded more privacy for more concentration of data collection with fewer tech companies. Stricter privacy rules may encourage even less transparency around app tracking, by shifting tracking code onto the servers of dominant tracking companies. Despite the new rules, large companies, like Google/Alphabet and Facebook/Meta, are still able to track users across apps, because these companies have access to unique amounts of first-party data about users. Apple is now able to track its customers even more accurately, by taking a larger share in advertising technologies and getting unique access to user identifiers, including the device serial number. This underlines that privacy and competition problems can be highly intertwined in digital markets and need holistic study.

Future work. In this Chapter, we only analysed apps that were already present on the App Store before iOS 14 and the ATT; it would be interesting to analyse how the ATT has impacted the privacy properties of *newly released* apps on the App Store, and how it might impact app privacy in the long run. It would also be helpful to develop a new automation tool for iOS apps to observe apps' data practices automatically, even beyond the first app start – as studied in this present Chapter – and thereby generate richer insights into these data practices. Furthermore, it would be pertinent to study user tracking by platforms in more detail, and also how the new Apple's Privacy Nutrition Labels inform the decision-making and education process of individuals around app privacy.

8

Discussion & Conclusions

Tracking has been known to be widespread in apps and highly intrusive for individuals. Despite this, there has been limited quantitative evidence of the current and changing nature of user choice over tracking. This is why this thesis aimed to provide robust quantitative insights.

8.1 Overview of Results

To analyse user choice over data within smartphones, this thesis traversed different levels at which users could potentially have a choice over tracking: regulatory interventions and electoral decisions (analysing the impact of the GDPR in Chapter 4), developer interventions and choice of apps (analysing the presence of consent in apps in Chapter 5), and platform interventions and choice of app platform (analysing the differences between Android and iOS and the impact of Apple's ATT in Chapters 6–7). Throughout, we considered the intersection between user choice over tracking and apps' compliance with important aspects of data protection law (e.g. the provision of consent flows). The assessment of compliance – and developing robust research tools for studying this in the first place – has been an important gap in previous literature.

8.1.1 Impact of the GDPR

In Chapter 4, we analysed the presence of tracking in apps, before and after the introduction of the GDPR. The aim of this is to gain insights into the impact of this data protection law on app tracking. We analysed 1m Android apps that our research group had previously downloaded in 2017, as well as another newly downloaded 1m Android apps from 2020 (see Section 3.1.1). We used similar analysis techniques to the previous study by Binns et al. [4], analysing the presence of tracker hosts in the app code with the `X-Ray 2020` database (see Section 3.1.2). We additionally reproduced the work of Binns et al. [47], who analysed market concentration in tracking in 5,000 apps and websites, but at a larger scale.

Tracking has remained prevalent across a wide range of mobile apps and prominent in its reach of app users. The number of tracking companies in the average app on Google Play has stayed about the same between 2017 and 2020. The top destination countries have likewise stayed the same, as have the most prominent tracking companies – namely Alphabet/Google and Meta/Facebook. 85% of apps from 2017 could send data to Alphabet/Google, compared to 89% in 2020. Apps still widely send personal data to tracking companies based in a third-party country without an 'adequate' level of data protection, particularly to those in the US. Our observations are consistent across 'super genres'. We found that the concentration in the tracking ecosystem has seen limited change over time. Competition between tracking companies seems to revolve at least partly around user privacy due to the relevance of little-known tracking companies. These can evade public and regulatory scrutiny to an extent, but still collect data about sizeable numbers of individuals. Of 53 observed M&A transactions in the tracking ecosystem between 2018 and 2020, only three were filed with EU or UK competition authorities.

We suspect that the observed limited change in the presence of tracking in apps stems from the fact that the underlying business models have not changed. App monetisation continues to rely on freemium and advertising-driven models. Apple and Google take a significant share of the revenue generated from this business

model, and face conflicts of interest in taming data collection and generating revenues from their respective app ecosystems.

Overall, our study suggests that the current enforcement of data protection obligations does not yet achieve its intended ends. Increased intervention by regulators is warranted, in particular as regards M&A transactions that concern data businesses and stricter enforcement of existing data protection rules.

8.1.2 Choice over Tracking in Apps

In Chapter 5, we first studied from a legal perspective whether and when apps need to ask EU and UK users for consent before engaging in third-party tracking. While there existed academic pieces on the conditions for tracking on the web, the case of tracking in apps had not been thoroughly studied before. This is despite smartphones usually having access to other and arguably more sensitive information, such as persistent cross-app identifiers to track user activity. We then analysed a representative sample of 1,297 apps from the Google Play Store – a random subset of the same 1m apps from 2020 that we had analysed in the previous Chapter 4. We manually opened each app on a real phone and checked for the presence of *any* consent flows (but did not interact with the app further), while logging transmissions to known tracking companies with the TrackerControl app and the `X-Ray 2020` database. Lastly, we studied the online guidance of the 13 most popular tracking companies, and how these companies support app developers in implementing consent.

Our legal analysis highlighted that apps must – with almost no exceptions – ask for consent from EU and UK users before engaging in tracking. This is because the ePrivacy Directive, which exists alongside the GDPR, requires consent before accessing or storing information on a user's device, unless strictly necessary for the functioning of the app. Since tracking is usually not strictly necessary for apps and inherently relies on accessing or storing information (to collect data about individuals), consent is required.

Our empirical findings suggest that very few apps actually ask for consent prior to tracking consumers (less than 10%), while most apps (more than 70%) share data with a range of third-party companies before any user interaction. In our analysis, we did not determine what these tracking companies do with this collected data, or if it is, in fact, used to track users, because we as researchers have limited insights into the servers and further data processing of tracking companies.

An important reason for this lack of compliance is that existing guidance for app developers by tracking companies is often hard to find, poorly maintained and difficult to read, due to the inherent conflicts of interest of tracking companies in increasing data collection and getting integrated into apps while protecting user privacy. Despite UK data protection law usually requiring consent to tracking, we found that Google and other companies do not make this point clear in their online implementation guidance for app developers and thereby make compliance for app developers harder than necessary.

8.1.3 Choice between iOS and Android Apps

In Chapter 6, we compared privacy in Android and iOS apps. An understanding of the iOS ecosystem is important for an informed consumer choice around app platforms, but the last large-scale study on iOS app privacy had been done in 2013 by Agarwal and Hall [33]. For our analysis, we adopted a varied set of metrics to assess privacy, with a focus on commonly used privacy metrics as well as measures for compliance with important obligations under EU, UK and US privacy law (e.g. the provision of consent flows). Using code and network analysis, we analysed a total of 24k apps from across the Apple App and Google Play Store (12k apps each). We selected a random subset of those apps in our app dataset (see Section 3.1.1) that were released or updated since January 2018. We made this choice to focus on apps that are currently in use. Apps were downloaded from the UK app stores around the beginning of 2020. This was about a year *before* Apple introduced new iOS privacy measures that may have led to changes in app tracking (see Chapter 7 for a follow-up investigation).

In our analysis, we found no clear winner in terms of privacy between iOS and Android across the various dimensions studied. Android apps tended to share an Advertising Identifier (known as 'Android Advertising Identifier' on Android and as 'Identifier for Advertisers' on iOS), which can be used for tracking users across apps, more often than iOS apps. Permissions, that both Apple and Google deem as particularly dangerous and require user opt-in, were more common among iOS apps (although Android also has a greater range of permissions that are deemed 'not dangerous' and do not require opt-in). On both platforms, our study highlights widespread potential violations of US, EU and UK data protection and privacy laws, including 1) the use of third-party tracking without user consent, 2) the lack of parental consent before sharing personal data with third-parties in children's apps, 3) the non-data-minimising configuration of tracking libraries, and 4) the sending of personal data to countries without an adequate level of data protection. More generally, we observed an absence of transparency around tracking, partly due to design decisions by Apple and Google. Such transparency is essential in keeping and holding gatekeeper power to account, but the analysis thereof remains difficult in the mobile tracking ecosystem. This conflicts with the strict transparency requirements for the processing of personal data laid out in the GDPR (Article 5).

8.1.4 Apple's Intervention against Tracking

Acknowledging the invasiveness of third-party tracking, Apple introduced two new privacy measures with iOS 14: Privacy Nutrition Labels and App Tracking Transparency (ATT). To assess the impact of these changes on app tracking, we analysed 1,759 iOS apps from the UK Apple App Store in Chapter 7: one version from before iOS 14 and one that has been updated to comply with Apple's new rules. We selected these apps by revisiting the same 12k iOS apps from the previous Chapter 6, and attempting to re-download all these apps (and then only including those apps that were compiled for iOS 14.5 or higher).

We found that Apple's new policies, as promised, prevent the collection of the Identifier for Advertisers (IDFA), an identifier used to facilitate cross-app user

tracking. However, the number of tracking libraries has – on average – roughly stayed the same in the studied apps. The average number of contacted companies and domains as well as the integrated opt-in permissions of iOS apps have seen a slight, but statistically significant, increase. We also found that the Privacy Nutrition Labels can be inaccurate and mislead consumers about apps' actual privacy practices. For example, 80.2% of those apps, that declared not to collect any data in their Privacy Nutrition Labels, actually sent data to at least one known tracking company right at the app's first initiation, before any user interaction and thus without user consent. This observed phenomenon seems to be more widespread in apps that did not range among the top apps on the App Store. As before, we did not analyse the invasiveness of such tracking because we do not have insights into the further data processing of tracking companies, so some of this observed data sharing with tracking companies may not be problematic.

Overall, Apple's technical changes make tracking more difficult now, but also reinforce the market power of existing gatekeeper companies with access to large troves of first-party data. Smaller data brokers, who used to engage in some of the most invasive data practices, will now face much higher challenges in conducting their business – a positive development for the privacy protections of end-users. We expect, however, that tracking companies will eventually work around these new policies, by using statistical methods ('fingerprinting') to identify users. Such fingerprinting would likely be easier to conduct for larger companies than smaller ones – deepening current imbalances in market power. A recent report by the Financial Times confirms this, and also highlights that Apple might foresee ways for other large companies to get around the ATT rules – something that might be unexpected for consumers [245]. Despite the new rules, large companies, like Google/Alphabet and Facebook/Meta, are still able to track users across apps, because these companies have access to unique amounts of data about users. Out of similar concerns, the CMA is investigating Google's Privacy Sandbox, which would entail the removal of third-party cookies from its Google Chrome browser [248].

In our analysis, we found convincing evidence that sophisticated fingerprinting methods are already used in practice.

In our study, we also observed that Apple still has access to a wide range of device identifiers that third-party app developers have no access to. For example, we observed that Apple regularly collects the device's serial number (see Figure 5a), through which Apple can accurately tie the point-of-sale of its devices to activities on the device itself, and track the device lifecycle in great detail. This is information that competitors do not have access to, and might disproportionately privilege Apple's position in the smartphone (data) market.

8.2 Reflections on Methodology

As outlined in Chapter 3, the methodology of this thesis had four main ingredients: a large app dataset of 2.3m apps (1m Android apps from 2017, 1m Android apps from 2020, 0.3m iOS apps from 2020), a new large-scale analysis method for iOS and Android (PlatformControl), the `X-Ray 2020` database and the TrackerControl app.

These approaches also have limitations. Due to the scale of apps analysed, we could not dissect apps individually, and there might be inaccuracies in our results. Less focus was put on the top apps by some of the most well-resourced app developers, since these developers are known to struggle less with compliance and also because the composition of these top apps is regularly changing. We only considered the data practices that we can study on-device; what happens behind the scenes, on the servers of tracking companies and between tracking companies, is not covered by our analysis. The lack of compliance on the client-side suggests that similar levels of compliance might be found behind the scenes, on the servers of tracking companies; the recent ruling by the Belgian data protection authority on the IAB's Transparency & Consent Framework confirms this [89, 249]. Some of our analysis, particularly the analysis of consent in apps, was semi-automated, and might be improved further through automation. The analysis of compliance remains difficult, because there exists no such thing as a single privacy score that covers all aspects

of app compliance. This is why, rather than focus on one single aspect as some previous work did, we drew on a range of important compliance aspects, including the provision of consent and the sending of personal data to non-'adequate' countries.

One of the most important aspects of the work in this dissertation is that the methods and data are public (at `https://www.platformcontrol.org/`) and can be reproduced easily. This supports keeping up with the quickly changing app ecosystem. Previous analysis of iOS apps relied on the decryption of apps, which is a legal grey area since it involves the circumvention of copyright protections (see Section 6.1.1). This acts as a deterrent to the development and publication of iOS privacy analysis tools. We managed to develop a new, `Frida`-based method that enables the analysis of iOS app privacy, without the need for decryption. Our approach is not perfect, since it currently only considers what libraries are present in apps and does not scale as well as commonly used Android app analysis tools (though the performance is similar to previous iOS analysis tools). There might be room for further improvement.

Ultimately, we hope that this dissertation will pave the way towards a new transparency and accountability instrument for apps' data practices.

8.3 Revisiting What Success Looks Like

8.3.1 A Tracking-Free Mobile Ecosystem?

Over recent months, the pressure on tracking providers has been increasing. In February 2022, the Belgian data protection authority found that the IAB's Transparency & Consent Framework is in violation of EU/UK data protection law [89, 249]. Among other aspects, the authority argued that the IAB is in fact a data controller and not just those organisations that use the IAB framework. This ruling by the authority represents one of the first applications of the EJC's rulings on joint controllership within the context of web and mobile tracking [117, 194, 195]. The ruling underlines that those who design the technical infrastructure behind the tracking ecosystem bear responsibility for their design decisions under EU/UK data protection law.

Around the same time, the Austrian and French data protection authorities as well as the European Data Protection Supervisor found that the use of Google Analytics on websites can be in violation of the ECJ's prohibition of personal data flows to the US without sufficient safeguards (*Schrems II ruling*) [60, 250–252]. These rulings suggest that the widespread sending of personal data to the US – which this dissertation proved to be common in app tracking (see Chapters 4 and 6) – faces an uncertain future. Without a new, reliable regime for the transatlantic sharing of personal data, the current practice of tracking is unlikely to be sustainable for much longer for app developers. The processing of personal data will need to find ways to overcome the reliance on US-centred infrastructure (and the potential harm posed by US intelligence agencies accessing these servers). Whether the proposed EU-U.S. Data Privacy Framework from 2022 is fit for the task remains to be seen.

As discussed in Chapter 5, the use of tracking relies on the fact that many app developers need ads for monetisation. Conversely, ads are the primary reason for using invasive tracking technologies. However, the invasiveness of tracking as well as recent rulings by European data protection authorities and courts cast doubt over the current practice. While ads and the personalisation thereof are often permissible, the use of invasive third-party tracking to support these technologies is often not. As a result, the link between personalised ads and tracking will likely weaken in the near future.

The industry is reacting to these recent developments and is working towards privacy-preserving advertising solutions. Apple and Google are increasingly preventing apps – and thereby third-party tracking companies – from accessing persistent user identifiers. Prominent recent examples are the introduction of the App Tracking Transparency framework on iOS (blocking access to unique user identifiers without user consent), the planned ban of third-party cookies from the Google Chrome browser (preventing websites from saving unique identifiers in cookies to track users across websites), and Google's introduction of a user opt-out from sharing personal identifiers with apps on Android. While these measures can increase

consumer privacy, they might also put more power over user data into the hands of the digital gatekeepers.

Increased restrictions on user identifiers might shift the tracking ecosystem in the direction of statistical identifiers (e.g. device fingerprinting). A company might then (likely wrongly) argue that these statistical identifiers do not fall under the protections of the GDPR anymore, since data cannot be uniquely attributed to an individual or only with great effort [253]. This argument is already used by the industry[5] to justify the use of pseudonymous identifiers, which, however, fall under the GDPR [255, 256]. While the threshold for the GDPR not to apply is high [253, 257], the increased use of statistical identifiers could make it more difficult for individuals to enjoy and exert their data protection rights in practice. At the same time, statistical identifiers may simply not be a good enough replacement for persistent user identifiers (such as advertising identifiers). If tracking systems do not have access to persistent user identifiers anymore, this might not only inhibit data trading, but might also make some smaller tracking companies less viable and run out of business.

One key technology for the mobile advertising industry is *install attribution*. When app A shows an ad by an advertising company to install app B, then this advertising company would like to know if a user has installed app B after clicking app A's ad ('conversion'). Traditionally, advertisers monitored conversions through a persistent user identifier, such as the IDFA. Now that Apple is significantly restricting access to the IDFA (through ATT) the company has implemented a new privacy-preserving ad attribution framework: SKAdNetwork. This new framework operates without persistent user identifiers, and discloses much less information about the user to advertisers; at the same time, Apple now gains more insights into the conversions of other advertisers. Indeed, Apple operates its own attribution framework that is not subject to the new ATT rules and gives advertisers much better insights into conversions.

[5]For example, Google argues that 'pseudonymous cookie IDs', 'pseudonymous advertising IDs', 'IP addresses', and 'other pseudonymous end user identifiers' do not fall under its own definition of 'Personally Identifiable Information' (PII) [254].

The case of SKAdNetwork shows that it is possible to build more privacy-preserving advertising technologies. It also shows that there is a risk that the shift towards more privacy will reinforce gatekeeper power and reduce competition around mobile ads. More competition around ads is usually a good thing for consumers because it will reduce the cost of ads and, as a result, the price of products they buy. Currently, we also have competition around tracking (surveillance) of users, since tracking still underpins a lot of online advertising. The result of competition around tracking has traditionally been rather negative for consumers because it creates a race to build better profiles and to collect more data about individuals, thereby conflicting with their data protection and privacy rights, among others.

It might sometimes seem difficult to imagine how the genie could be put back in the bottle in the tracking ecosystem, given that the app ecosystem relies on the income generated from tracking. However, this is what Apple is currently attempting – phasing out mobile tracking over time and shifting towards more privacy-preserving advertising technologies. This makes technical changes to the tracking ecosystem more important than ever and will need further scrutiny in the future.

8.3.2 Alternative App Stores on iOS?

There is a long-running debate about whether alternative iOS app stores might improve the conditions for consumers. These proposals have even made it into the EU Digital Markets Act from 2022. This will likely force Apple to allow users to install apps from outside the official Apple App Store.

Historically, alternative app stores have not fared well. Well-resourced companies such as Blackberry and Amazon have attempted to create alternative app stores within the Android ecosystem. Blackberry implemented an Android subsystem into its Blackberry OS 10 that allowed users to install Android apps alongside their Blackberry apps. The company eventually withdrew from the smartphone business. Amazon launched its Appstore in 2011, as an alternative to Google's system, but has never seen wide adoption.

The closed environments of both the Google and Apple app stores come with a range of benefits. They offer good usability to end-users because they serve as a single point of contact. They also foster user trust and security. It would have been arguably more difficult for Apple to enforce its new ATT policies to reduce the invasiveness of tracking without being the provider of the Apple App Store as well.

Even if alternative app stores will not be widely adopted, they still show some promise and can serve as an important testing ground for new ideas. For example, Amazon pioneered subscription models for paid apps. Both Apple and Google now have similar models. Similarly, the Aurora app store was among the first to show privacy nutrition labels on its store, which are now implemented by Google and Apple in their own app stores as well.

The often-used argument of Apple that sideloading ('a cybercriminal's best friend' [258]) would undermine the security of its iOS devices is rather weak. Sideloading has long been possible on Android, and Google has implemented a range of measures to safeguard this process. Play Protect – by default – transmits all sideloaded apps to Google for safety checks. Android also implements a careful UI design to avoid less experienced users from sideloading. Arguing that sideloading is inherently insecure would imply that the implementation of Android has been deeply flawed since its inception more than a decade ago. The argument also discredits the merits of openness, transparency and interoperability within computer systems, and might be seen as overly paternalistic by many iOS users. Apple's refusal of sideloading also makes it harder for researchers to access much-needed insights into app privacy within mobile ecosystems, as discussed in the following.

8.3.3 Easy Access to Insights for Researchers

On both Android and iOS, the download and analysis of apps at scale remains challenging for researchers. This is, however, essential to app privacy and security research. As previously discussed, Apple applies encryption to all iOS apps by default, which drives researchers into legal grey areas (see Section 6.1.1). As a result, there has been hardly any large-scale research into the privacy practices

of iOS apps from 2013 [33] until the publication of our 2022 paper on comparing iOS and Android privacy (Chapter 6). We have worked around some, but by far not all, the limitations in analysing iOS apps.

As discussed in Section 5.1, Google has been introducing various measures that are meant to increase user privacy and security on Android, but have also made research significantly more difficult. The most notable of such measures is the *ban on installing self-signed certificates* (thereby preventing researchers from analysing apps' network traffic for app research without deep modifications of the system files of Android devices) and the rollout introduction of the Google SafetyNet (which makes it impossible to run certain apps – including popular apps like Snapchat and Pokémon Go – on Android devices with modified system files). The rollout of the SafetyNet and the ban on self-signed certificates in tandem makes app research like ours extremely difficult. While Google argued that the ban on self-signed certificates would serve device security, it seems that the company could easily implement choice architectures for average end-users to prevent them from accidentally installing such certificates (as is currently done on iOS where the installing of such certificates is not easy but possible), while still allowing researchers to disable such security features to conduct their work. Some internet outlets even declared the death of modifying the Android operating system (currently a central requirement for Android app research) in response to Google's rollout of the SafetyNet [169]. Additionally, many apps nowadays use code obfuscation (see Section 6.2.2.1), which further complicates app privacy research.

With our publicly available app download and analysis methods (available at `https://www.platformcontrol.org/`), we hope that researchers will be able to access insights into apps' data practices more easily in the future.

8.3.4 The Need for Better User Controls?

Some, particularly from within the tracking industry, argue for better user controls to tackle problems around data protection on the Internet. These voices have

arguably led to a proliferation of cookie banners, but did little to improve better protections of users' fundamental rights online [5, 7].

There exists ample research on the fact that privacy self-management is a flawed concept [25, 27, 44, 49, 52, 54]. Individuals by themselves struggle to navigate the vast number of privacy choices that they face in their online lives, and thus rely on the choice architectures presented by app developers and platforms. Users do often not understand the options provided and will follow their intuition when reacting to any prompt provided. This explains the high signup rates on the web, where users often are nudged into 'consent'. This also explains the high opt-out rates on iOS, which provides users with two rather equal options regarding whether to accept or decline app tracking [158, 159, 232]. This, in turn, underlines that user opt-out is often a matter of software *design*, rather than user *choice* [2, 5, 7].

At the same time, consent still forms a central part of data protection regimes worldwide, including the EU, UK and US. While individual users struggle to manage privacy, the sum of privacy preferences can indeed influence the regulatory environment (as seen by the introduction of the GDPR in 2016) and the technical implementation of choice architectures around data (as seen by recent privacy efforts of Google and Apple). These efforts face significant headwinds, not the least by the tracking industry itself which has long advocated its Transparency & Consent Framework as a legitimate solution to comply with the consent obligations under the GDPR. As mentioned earlier in this dissertation, there is a broad agreement by academics, regulators and activists that the status quo around (user choice over) tracking is not compliant with the GDPR [5, 7, 84, 89, 143, 259–261]. This has also been highlighted by the research in this dissertation.

This implies that, to improve the privacy protections of users, the choice architecture around data should be brought within the realm of the existing legal obligations. This does not necessarily mean better user controls, which are often ineffective, but a serious consideration of the concept of *data protection by design and default*, as is already foreseen by Article 25 of the GDPR. As regards the case of invasive tracking, there seems to exist limited space within the regulatory

environment to justify such practice *per se*, let alone adding any user controls around them. In the interim, more transparency and accountability, as aimed for by the tools developed in this dissertation, could increase the level of *meaningful* user choice, by encouraging more public debate and pressure around permissible business models and data practices in the digital age.

8.4 A New Approach to Tech Regulation

Given the current widespread mismatch between the law on the books and data practices in reality, iterative changes to current legal practice will not be enough. This is not a problem of the law *per se* but rather one of philosophy and established patterns of thought. Based on this and other research, we propose a range of priorities for future work to move beyond the status quo.

Priority 1: Make consent *meaningful* – or abandon it

Although often claimed otherwise, the GDPR *does not* require a broad implementation of 'cookie banners'. EU and UK data protection principles have hardly changed since the GDPR came into force in May 2018 and were already part of the Data Protection Directive 1995. The requirements regarding 'cookie banners' additionally result from Art 5(3) of the ePrivacy Directive as amended in 2009, not from the GDPR.

The recent flood of cookie banners can rather be explained by the fact that the potential sanctions for data protection violations have drastically increased with the GDPR, causing discontent within the online data industry. The conditions for consent have also been tightened and existing standards have been clarified in line with case law. User consent must now be 'freely given, specific, informed and unambiguous' (Recital 32 GDPR). However, this is rarely the case in practice, as found in this and other work. A significant proportion of current 'cookie banners' are thus in violation of the GDPR.

The designation of those *consent* banners as 'cookie banners' can further be interpreted as misinformation. For example, Facebook implements a pop-up on its website titled 'Allow the use of cookies from Facebook in this browser?' It is only in the accompanying Cookie Policy that Facebook clarifies that 'cookies' do not only refer to cookies but also that other 'technologies, including data that we store on your web browser or device, identifiers associated with your device and other software, are used for similar purposes.' The online advertising industry today uses a variety of technologies to track user activity across various apps and websites – such as fingerprinting (i.e. using browser characteristics such as time zone, language and operating system) and email hashing (i.e. sending email addresses from non-Facebook websites to Facebook even if the user does not use Facebook). This collection of data about websites and apps – tracking – is widespread, as found by this thesis and other research. Meanwhile, the term 'cookie' sounds innocuous and is widely used by the industry.

Overall, a considerable part of the 'cookie banners' on the internet aims to misinform and frustrate internet users vis-à-vis the GDPR rather than to implement the law's requirements [249].

There remains significant work to do for authorities and other organisations to tackle incompliant implementations of consent and make it meaningful. Indeed, ample research suggests that this is not possible at all [25–27], in part because individuals will never be sufficiently 'informed' – as is required by GDPR – about the opaque data practices of large technology companies [262].

Priority 2: Better, bolder communication

Due to the continued uncertainty and misinformation regarding the GDPR, the current way of working of data protection and other public authorities has created a vacuum of knowledge and authority that has been successfully occupied by third parties with strong self-interests. This way of working in the EU is often characterised as bureaucratic and apolitical, resulting from a lack of a transnational

public sphere in Europe. However, without a European public sphere and debate, political legitimacy in the Habermasian sense is difficult, if not impossible.

The end result is problematic for data protection because it fuels a negative and dismissive mood among citizens – including those individuals who are responsible for the practical implementation of the GDPR – towards the competence of public authorities in digital matters. Data protection and other public authorities should counter this perception boldly and decisively. This applies both to new digital initiatives and existing laws such as the GDPR.

Priority 3: Clear technical standards, visualisations and reference code

As part of better communication, *clear, reliable and actionable technical standards* should be considered. Unfortunately, developers do often not know how to comply, so there is a need to clarify what forms of data processing are permitted and how this should be implemented in software.

Currently, the expectation from the authorities is that software developers will resolve important issues related to the implementation of the GDPR themselves – by studying the relevant legislation and rulings. This assumption is unrealistic, at least for smaller software companies [32]. In addition, the European Data Protection Board and the ICO regularly publish explanatory notes on important aspects of the GDPR. This usually involves the publication of long texts of legalese. The target audience of these publications is thus primarily legal, especially courts, but not the individuals tasked with the practical implementation of the law.

It is certainly important to explain the legal dimensions of the GDPR and to pursue this through legal methodology, particularly by publishing explanatory legal texts. At the same time, it seems that authorities too often hide their lack of authority and technical expertise behind overly formal communication and shy away from clear specifications. As a result, a significant part of the interpretation of the GDPR currently falls to the courts. Unfortunately, this approach undermines a swift and effective implementation of the GDPR and is unsuitable to keep pace

with rapid technological change. Code can be changed and rolled out to users worldwide in a matter of minutes. For effective IT regulation, the (ambitious) goal must be to act similarly agile.

From a technical perspective, it is almost naïve to assume that legal text could be translated more or less directly into code. Instead, in IT, *requirements specification* provides a decades-old approach to describing and building IT systems. A common standard was first published by the IEEE (Institute of Electrical and Electronic Engineers) in 1984; the latest version is ISO/IEC/IEEE 29148:2018 from 2018. Requirements can be both technical and non-technical as well as specific and less specific. There is no reason why similar requirements cannot be formulated for core elements of the GDPR and other IT law. This could be done in particular for the implementation of consent and should be accompanied by visualisations and reference code where possible. In the context of the amendment of the ePrivacy Directive in 2009, the EU even provided visualisations and reference code in the past, but did not maintain them over the years and discontinued them after the introduction of the GDPR.

Priority 4: Sufficient resources for authorities

There are many reasons for the lack of implementation of the GDPR in practice. One important reason is the continued lack of resources of data protection authorities [263]. This refers to both financial resources and (technical) expertise. For example, there has been virtually no action by the responsible authorities against the documented data protection problems in mobile apps. A second reason is the one-stop-shop approach of the GDPR in the EU. This approach currently leads to a race to the bottom between member states in terms of negligent implementation of the GDPR. In particular, Ireland, where most of the major tech companies in Europe are based (including Microsoft, Alphabet/Google and Meta/Facebook), has been criticised in this regard [144, 146]. A third reason is the still-evolving case law in the courts. Since the GDPR is still relatively new, there are still many aspects of the law that are still being clarified by the courts.

The problem of the lack of practical enforcement of the GDPR has been recognised by lawmakers and is being addressed in new EU digital legislation. Ireland is no longer a single point of failure of legal enforcement in DMA and DSA against big tech, as it was de facto under the GDPR, but rather the EU Commission. In addition, technology companies will be required to subsidise enforcement financially.

A key challenge will remain recruiting the necessary technical talent for public institutions, most of whom currently work for the same technology companies and are needed to keep pace with private industry in terms of expertise and technical understanding. In the past, European legislators have not always maintained an air of technical competence. One example is the planned EU AI Act, which is supposed to regulate AI applications. However, the definition of AI applications in the Commission's first proposal was so broad that it covered almost any computer application. The planned AI rules are derived from EU product safety legislation. This creates the risk of missing the core of AI, which rather lies in the inputs and outputs of the model rather than the product/technology itself. Doubts about the EU legislator's deep technical understanding also arise when reading the GDPR. The law, like its 1995 predecessor, distinguishes between controllers and processors in the processing of personal data. Controllers are those that alone or jointly with others determine the purposes and means of the processing of personal data (Art 4 GDPR). Processors, on the other hand, usually only act at the instruction of the controller. However, today's IT systems are the product of the combined work of many different actors, both large and small. This often makes it almost impossible to distinguish between controllers and processors. This distinction is, however, important because controllers face many more obligations than processors. Moreover, whether and to what degree software development – rather than direct data processing – entails obligations under the GDPR is not clear [127]. As a result of these definitions, which only peripherally deal with the usual processes and distribution of tasks in software development, there are a number of concluded and pending cases regarding the definition of the role of data controller [89, 117, 194, 195]. One solution to this phenomenon was proposed by the Belgian DPA in

the aforementioned case against IAB Europe: the DPA decided to define almost all actors in the online advertising business as controllers, i.e. thousands of different companies. IAB Europe has appealed and the case is currently pending before the ECJ. As an alternative approach, China's Personal Information Protection Law (PIPL) from 2021 only foresees processors of personal data, but no controllers.

Of course, the GDPR is not limited to IT but covers many other areas of our daily lives that involve personal data. Therefore, one could argue that criticism of the GDPR's lack of focus on software development misses the point of the law. However, it is also the case that without technological developments, there would have been little motivation for a revision of EU data protection law (see also Recital 6 GDPR).

Priority 5: Embrace regulatory technologies

There are two dominant approaches to enforcing data protection rules in digital systems. The first one is taken by data protection authorities who tend to focus their efforts on a few select cases and companies. The hope is that this will tame the most excessive data practices and that there will be spillover effects across the data practices by other organisations. The second approach is taken by gatekeepers, such as app stores, who conduct some enforcement of data protection rules at scale (e.g. through their (automated) app review), but publish limited public information about this enforcement, including the *number and nature* of decisions taken. Given the scale of the digital ecosystem and the extent of current violations of data protection rules (as observed in this and other work), both approaches are insufficient. Without the help of regulatory technologies in ensuring compliance in the digital ecosystem, it will be *impossible* to scale operations across the vastness of these digital ecosystems, to fulfil the expectations of individuals in keeping them safe online, and to protect fundamental rights and freedoms.

In the app ecosystem, an important, persisting issue that emerged from our analysis across iOS and Android is the lack of transparency around apps' data practices. This conflicts with the strict transparency requirements for the processing of personal data laid out in the GDPR. Design decisions by Apple and Google

currently impede research efforts, such as the application of copyright protection to *every* iOS app – even free ones. I have elaborated on this aspect in a recent preprint [264]. This is why it is *important to develop and maintain transparency tools*, as undertaken in this research. Future work should strive to expand the capabilities of the presented PlatformControl toolkit, and give more up-to-date and detailed insights into apps' privacy and compliance properties. As part of this, an important field for further study is the development of a cross-platform app instrumentation tool. *Automatic app compliance analysis tools* are not widely available nor used by regulators and the interested public (though it might be easy to conduct such automatic checks if the regulators defined more explicit rules regarding privacy and app design), but would help keep up with the vastness of the app ecosystem. Such analysis tools would require *reliable and computable metrics for compliance*. While most of this work has been on the situation in Europe, there have been emerging many promising new pieces of technology regulation across the globe, which need further investigation.

Priority 6: Evolve 'legacy' legislation and provide support for research

This thesis devoted much of its efforts to analysing GDPR compliance in mobile apps. Such research is currently challenging, as the creation of the necessary data is associated with high investments of time and scarce technical expertise. The fact that analysing privacy issues in apps and in other software products is so difficult has an impact not only on our research but also on the work of other researchers and data protection authorities aiming to protect fundamental rights in digital systems. For example, most data protection authorities themselves currently do not possess independent expertise to analyse compliance issues in mobile apps.

The EU Digital Services Act makes promising progress in supporting research in relation to online platforms and search engines. Its Article 40, for example, obliges 'very large online platforms' and 'very large online search engines' to allow researchers to analyse 'systemic risks'. The concrete implications for research

practice, however, remain to be seen. I have explored this aspect in a recent preprint [264]. It can, however, be expected that clarifications of the law by the highest courts will be necessary and that many years will pass before the law will lead to major changes to the status quo.

Despite all the debates about new IT laws, one must not lose sight of existing laws, such as copyright, patent, and IT security law. Even if such legislation may be less attractive for public and academic debate and thus receives less attention, there is also a great need for improvement here. This was also demonstrated by the research in this thesis, which avoided challenges related to copyright law and Apple's use of DRM in iOS apps.

8.5 Conclusions

This thesis analysed app tracking at four levels: the impact of the GDPR (Chapter 4), consent to tracking in apps (Chapter 5), differences between Android and iOS (Chapters 6), and the impact of Apple's App Tracking Transparency (ATT) framework (Chapter 7). While many previous studies looked at data protection and privacy in apps, few studies analysed tracking over time, took a compliance angle, or looked at iOS apps at scale. Throughout our analysis of apps, we found compliance problems within apps as regards key aspects of US, EU and UK data protection and privacy law, particularly the need to seek consent before tracking. For instance, while user consent is usually required prior to tracking in the EU and UK (under the ePrivacy Directive), our empirical findings suggest that tracking takes place widely and usually without users' awareness or explicit agreement.

This thesis emphasised the need for robust analysis methods to build transparency and accountability around app privacy and the actions of all relevant stakeholders, including platforms, tracking companies, developers and regulators. To pave the way towards this goal, this thesis contributed 1) a scalable downloading and analysis framework for iOS and Android privacy and compliance analysis (PlatformControl), 2) an improved understanding of the legal requirements and

empirical facts regarding app tracking, 3) a comprehensive database of the relations between companies in the app ecosystem (X-Ray 2020), and 4) an Android app to support the easy and independent analysis of apps' privacy practices (TrackerControl). The app download and analysis tools are publicly available at https://www.platformcontrol.org/.

Privacy is still broken in the app ecosystem, and there will be no single fix. It will be essential for any fix that we know the facts on the ground. In the meantime, there will continue to be a widespread sense within the tracking ecosystem that it is fine to extract prodigious amounts of data from individuals, without their genuine awareness or control. This can have significant effects on the rights to privacy and data protection, but also on other fundamental rights, such as the right to non-discrimination (e.g. when data from mobile tracking is used in AI systems, such as targeted ads for job offers) or the right to free and fair elections (e.g. when political microtargeting is used, as in the Brexit vote or the Trump election). Crucially, user choice – a key theme of this thesis – does not necessarily need to be implemented as consent. If data protection laws were adequate, sufficiently enforced and supported by citizens, this might already ensure a reasonable level of data protection.

Ultimately, this thesis concludes that it is difficult to see how ubiquitous surveillance of individuals could ever be compatible with the promises and foundations of Western liberal democracies. *If* it is indeed incompatible, the choice that individuals should have over tracking is none at all, and these practices should have no place in our society.

References

[1] A. Razaghpanah, R. Nithyanand, N. Vallina-Rodriguez, S. Sundaresan, M. Allman, C. Kreibich and P. Gill. 'Apps, Trackers, Privacy, and Regulators: A Global Study of the Mobile Tracking Ecosystem'. In: *Proceedings 2018 Network and Distributed System Security Symposium*. San Diego, CA: Internet Society, 2018. ISBN: 978-1-891562-49-5. DOI: 10.14722/ndss.2018.23353.

[2] M. Van Kleek, I. Liccardi, R. Binns, J. Zhao, D. J. Weitzner and N. Shadbolt. 'Better the Devil You Know: Exposing the Data Sharing Practices of Smartphone Apps'. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. Chi '17. Denver, Colorado, USA: ACM Press, May 2017, pp. 5208–5220. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025556.

[3] S. Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. 1st ed. 2019. ISBN: 978-1781256855.

[4] R. Binns, U. Lyngs, M. Van Kleek, J. Zhao, T. Libert and N. Shadbolt. 'Third Party Tracking in the Mobile Ecosystem'. In: *Proceedings of the 10th ACM Conference on Web Science - WebSci '18*. the 10th ACM Conference. WebSci '18. New York, NY, United States: ACM Press, May 2018, pp. 23–31. ISBN: 978-1-4503-5563-6. DOI: 10.1145/3201064.3201089.

[5] M. Nouwens, I. Liccardi, M. Veale, D. Karger and L. Kagal. 'Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence'. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Chi '20. New York, NY, USA: Association for Computing Machinery, 2020. DOI: 10.1145/3313831.3376321.

[6] K. Kollnig, R. Binns, P. Dewitte, M. Van Kleek, G. Wang, D. Omeiza, H. Webb and N. 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)*. USENIX Association, Aug. 2021, pp. 181–196. ISBN: 978-1-939133-25-0. URL: https://www.usenix.org/conference/soups2021/presentation/kollnig.

[7] C. Matte, N. Bielova and C. Santos. 'Do Cookie Banners Respect my Choice? Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework'. In: *2020 IEEE Symposium on Security and Privacy (SP)* (2019), pp. 791–809. DOI: 10.1109/sp40000.2020.00076.

[8] I. Shklovski, S. D. Mainwaring, H. H. Skúladóttir and H. Borgthorsson. 'Leakiness and Creepiness in App Space: Perceptions of Privacy and Mobile App Use'. In: *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*. The 32nd Annual ACM Conference. Chi '14. New York, NY, United States: ACM Press, 2014, pp. 2347–2356. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557421.

[9] A. M. McDonald and L. F. Cranor. 'The Cost of Reading Privacy Policies'. In: *I/S: A Journal of Law and Policy for the Information Society* (2008). URL: https://lorrie.cranor.org/pubs/readingPolicyCost-authorDraft.pdf.

[10] S. Vosoughi, D. Roy and S. Aral. 'The Spread of True and False News Online'. In: *Science* 359.6380 (2018), pp. 1146–1151. DOI: 10.1126/science.aap9559.

[11] Centre for Data Ethics and Innovation. *Review of Online Targeting: Final Report and Recommendations*. 2020. URL: https://www.gov.uk/government/publications/cdei-review-of-online-targeting/online-targeting-final-report-and-recommendations.

[12] M. House of Commons: Digital Culture and S. Committee. *Disinformation and 'fake news': Final Report*. Feb. 2019. URL: https://www.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/news/fake-news-report-published-17-19/.

[13] F. M. Shipman and C. C. Marshall. 'Ownership, Privacy, and Control in the Wake of Cambridge Analytica: The Relationship between Attitudes and Awareness'. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: ACM, 2020. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376662.

[14] S. Wachter. 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising'. In: *Berkeley Technology Law Journal* 35.2 (2019), pp. 367–430. DOI: 10.15779/z38js9h82m.

[15] R. Benjamin. *Race after Technology: Abolitionist Tools for the New Jim Code*. Cambridge, United Kingdom: Polity, 2019. 1 p. ISBN: 978-1-5095-2643-7.

[16] U. Lyngs, K. Lukoff, P. Slovak, W. Seymour, H. Webb, M. Jirotka, J. Zhao, M. Van Kleek and N. Shadbolt. ''I Just Want to Hack Myself to Not Get Distracted': Evaluating Design Interventions for Self-Control on Facebook'. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20: CHI Conference on Human Factors in Computing Systems. Honolulu HI USA: Acm, 2020. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376672.

[17] G. Kovacs, D. M. Gregory, Z. Ma, Z. Wu, G. Emami, J. Ray and M. S. Bernstein. 'Conservation of Procrastination: Do Productivity Interventions Save Time Or Just Redistribute It?' In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. Glasgow, Scotland Uk: ACM Press, 2019. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300560.

[18] J. Colnago, Y. Feng, T. Palanivel, S. Pearman, M. Ung, A. Acquisti, L. F. Cranor and N. Sadeh. 'Informing the Design of a Personalized Privacy Assistant for the Internet of Things'. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu HI USA: Acm, Apr. 2020. ISBN: 978-1-4503-6708-0. DOI: 10.1145/3313831.3376389.

[19] Bundeskartellamt. *B6-22/16 (Facebook v Bundeskartellamt)*. 2019. URL: http://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Entscheidungen/Missbrauchsaufsicht/2019/B6-22-16.pdf%5C%3F__blob%5C%3DpublicationFile%5C%26v%5C%3D5.

[20] Digital Advertising Alliance. *YourAdChoices.Com*. 2023. URL: https://youradchoices.com/.

[21] IAB Europe. *TCF – Transparency & Consent Framework*. 2023. URL: https://iabeurope.eu/transparency-consent-framework/.

[22] W3C Working Group. *Tracking Preference Expression (DNT)*. 2019. URL: https://w3c.github.io/dnt/drafts/tracking-dnt.html.

[23] Global Privacy Control Consortium. *Global Privacy Control — Take Control Of Your Privacy*. 2023. URL: https://globalprivacycontrol.org/.

[24] Apple. *User Privacy and Data Use*. 2021. URL: https://developer.apple.com/app-store/user-privacy-and-data-use/.

[25] D. J. Solove. 'Privacy Self-Management and the Consent Dilemma'. In: *Harvard Law Review* 126 (2012). URL: https://harvardlawreview.org/wp-content/uploads/pdfs/vol126_solove.pdf.

[26] S. Barocas and H. Nissenbaum. 'On notice: The trouble with notice and consent'. In: *Proceedings of the engaging data forum: The first international forum on the application and management of personal electronic information*. 2009. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2567409.

[27] E. Bietti. 'Consent as a Free Pass: Platform Power and the Limits of the Informational Turn'. In: *Pace Law Review* (2020). URL: https://digitalcommons.pace.edu/cgi/viewcontent.cgi?article=2013&context=plr.

[28] L. Lessig. *Code 2.0*. 1st ed. Basic Books, 2006. ISBN: 978-0-465-03914-2.

[29] Giovanni Buttarelli. *The EU GDPR as a Clarion Call for a New Global Digital Gold Standard | European Data Protection Supervisor*. 2016. URL: https://edps.europa.eu/press-publications/press-news/blog/eu-gdpr-clarion-call-new-global-digital-gold-standard.

[30] A. Ekambaranathan, J. Zhao and M. Van Kleek. '"Money makes the world go around": Identifying Barriers to Better Privacy in Children's Apps From Developers' Perspectives'. In: *Conference on Human Factors in Computing Systems (CHI '21)*. The 2021 CHI Conference. ACM Press, 2021. DOI: 10.1145/3411764.3445599.

[31] A. H. Mhaidli, Y. Zou and F. Schaub. '"We Can't Live Without Them!" App Developers' Adoption of Ad Networks and Their Considerations of Consumer Risks'. In: *Proceedings of the Fifteenth Symposium on Usable Privacy and Security* (2019).

[32] S. Sirur, J. R. Nurse and H. Webb. 'Are We There Yet?: Understanding the Challenges Faced in Complying with the General Data Protection Regulation (GDPR)'. In: *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security - MPS '18*. the 2nd International Workshop. ACM Press, 2018, pp. 88–95. ISBN: 978-1-4503-5988-7. DOI: 10.1145/3267357.3267368.

[33] Y. Agarwal and M. Hall. 'ProtectMyPrivacy: Detecting and Mitigating Privacy Leaks on iOS Devices Using Crowdsourcing'. In: *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services - MobiSys '13*. Proceeding of the 11th Annual International Conference. MobiSys '13. Taipei, Taiwan: ACM Press, 2013. ISBN: 978-1-4503-1672-9. DOI: 10.1145/2462456.2464460.

[34] K. Kollnig and N. Shadbolt. 'TrackerControl: Transparency and Choice around App Tracking'. In: *Journal of Open Source Software* 7.75 (2022). DOI: 10.21105/joss.04270.

[35] K. Kollnig, R. Binns, M. Van Kleek, U. Lyngs, J. Zhao, C. Tinsman and N. Shadbolt. 'Before and after GDPR: Tracking in Mobile Apps'. In: *Internet Policy Review* 10.4 (2021). DOI: 10.14763/2021.4.1611.

[36] K. Kollnig, A. Shuba, R. Binns, M. Van Kleek and N. Shadbolt. 'Are iPhones Really Better for Privacy? A Comparative Study of iOS and Android Apps'. In: *Proceedings on Privacy Enhancing Technologies* 2022.2 (2022), pp. 6–24. DOI: 10.2478/popets-2022-0033.

[37] K. Kollnig, A. Shuba, M. Van Kleek, R. Binns and N. Shadbolt. 'Goodbye Tracking? Impact of IOS App Tracking Transparency and Privacy Labels'. In: *2022 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 508–520. DOI: 10.1145/3531146.3533116.

[38] K. Kollnig. 'Lehren Aus Der DSGVO: Fünf Prioritäten Für Wirksame IT-Regulierung'. In: *Ad Legendum* 2023.2 (2023).

[39] Y. Benkler. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press, 2006. 515 pp. ISBN: 978-0-300-11056-2.

[40] O. Lynskey. *The Foundations of EU Data Protection Law*. 1st ed. Oxford Studies in European Law. Oxford University Press, 2015. ISBN: 978-0-19-871823-9.

[41] L. Floridi. 'The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU'. In: *Philosophy & Technology* 33.3 (2020), pp. 369–378. DOI: 10.1007/s13347-020-00423-6.

[42] J. Reidenberg, J. Bhatia, T. D Breaux and T. B Norton. 'Ambiguity in Privacy Policies and the Impact of Regulation'. In: *The Journal of Legal Studies* 45.S2 (June 2016), S163–s190. DOI: 10.1086/688669.

[43] A. Acquisti. 'Nudging Privacy: The Behavioral Economics of Personal Information'. In: *IEEE Security & Privacy Magazine* 7.6 (2009), pp. 82–85. DOI: 10.1109/msp.2009.163.

[44] P. A. Norberg, D. R. Horne and D. A. Horne. 'The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors'. In: *Journal of Consumer Affairs* 41.1 (2017), pp. 100–126. DOI: 10.1111/j.1745-6606.2006.00070.x.

[45] D. J. Solove. 'The myth of the privacy paradox'. In: *George Washington Law Review* 89.1 (2021). URL: https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2738&context=faculty_publications.

[46] A. E. Waldman. 'Cognitive biases, dark patterns, and the 'privacy paradox''. In: *Current opinion in psychology* 31 (2020), pp. 105–109. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3456155.

[47] R. Binns, J. Zhao, M. Van Kleek and N. Shadbolt. 'Measuring Third-party Tracker Power across Web and Mobile'. In: *ACM Transactions on Internet Technology* 18.4 (2018). DOI: 10.1145/3176246.

[48] Competition and Markets Authority. *Online Platforms and Digital Advertising.* Market study final report. 2020. URL: https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final%5C_report%5C_1%5C_July%5C_2020%5C_.pdf.

[49] S. Wachter and B. Mittelstadt. 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI'. In: *Columbia Business Law Review* 2019.2 (2019). DOI: 10.31228/osf.io/mu2kf.

[50] J. P. Choi, D.-S. Jeon and B.-C. Kim. 'Privacy and personal data collection with information externalities'. In: *Journal of Public Economics* 173 (2019), pp. 113–124. DOI: 10.1016/j.jpubeco.2019.02.001.

[51] D. Bergemann, A. Bonatti and T. Gan. 'The Economics of Social Data'. In: *Cowles Foundation Discussion Papers* (2020). arXiv: 2004.03107.

[52] H. Nissenbaum. 'Privacy as Contextual Integrity'. In: *Washington Law Review* 79 (2004). URL: https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10/.

[53] A. Acquisti, L. Brandimarte and G. Loewenstein. 'Privacy and Human Behavior in the Age of Information'. In: *Science* 347.6221 (2015), pp. 509–514. DOI: 10.1126/science.aaa1465.

[54] A. Acquisti, C. Taylor and L. Wagman. 'The Economics of Privacy'. In: *Journal of Economic Literature* (2016). DOI: 10.1257/jel.54.2.442.

[55] G. M. Greco and L. Floridi. 'The Tragedy of the Digital Commons'. In: *Ethics and Information Technology* 6.2 (2004), pp. 73–81. DOI: 10.1007/s10676-004-2895-2.

[56] European Commission. *Flash Eurobarometer 443: Report e-Privacy.* 2016. URL: https://data.europa.eu/data/datasets/s2124%5C_443%5C_eng?locale=en.

[57] L. Dencik and J. Cable. 'The advent of surveillance realism: Public opinion and activist responses to the Snowden leaks'. In: *International Journal of Communication* 11 (2017), pp. 763–781. URL: https://ijoc.org/index.php/ijoc/article/view/5524.

[58] S. Zimmeck, P. Story, D. Smullen, A. Ravichander, Z. Wang, J. Reidenberg, N. C. Russell and N. Sadeh. 'MAPS: Scaling Privacy Compliance Analysis to a Million Apps'. In: *Proceedings on Privacy Enhancing Technologies* 2019.3 (June 2019), pp. 66–86. DOI: 10.2478/popets-2019-0037.

[59] H. Wang, Z. Liu, J. Liang, N. Vallina-Rodriguez, Y. Guo, L. Li, J. Tapiador, J. Cao and G. Xu. 'Beyond Google Play: A Large-Scale Comparative Study of Chinese Android App Markets'. In: *Proceedings of the Internet Measurement Conference 2018.* Imc '18. 2018, pp. 293–307. ISBN: 978-1-4503-5619-0. DOI: 10.1145/3278532.3278558.

[60] Court of Justice of the European Union. *Data Protection Commissioner v Facebook Ireland and Maximillian Schrems*. 2020. URL: https://curia.europa.eu/juris/documents.jsf?num=C-311/18.

[61] Apple. *Apple Advertising & Privacy*. 2021. URL: https://www.apple.com/legal/privacy/data/en/apple-advertising/.

[62] microg. *Implementation Status*. 2020. URL: https://github.com/microg/GmsCore/wiki/Implementation-Status.

[63] J. Gamba, M. Rashed, A. Razaghpanah, J. Tapiador and N. Vallina-Rodriguez. 'An Analysis of Pre-installed Android Software'. In: *41st IEEE Symposium on Security and Privacy*. May 2020. DOI: 10.1109/sp40000.2020.00013.

[64] Commission Nationale de l'Informatique et des Libertés. *Délibération SAN-2019-001 du 21 janvier 2019*. 2019. URL: https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000038032552/.

[65] D. J. Leith. 'Mobile Handset Privacy: Measuring the Data iOS and Android Send to Apple and Google'. In: *Security and Privacy in Communication Networks*. Ed. by J. Garcia-Alfaro, S. Li, R. Poovendran, H. Debar and M. Yung. Cham: Springer International Publishing, 2021, pp. 231–251. ISBN: 978-3-030-90022-9. URL: https://www.scss.tcd.ie/doug.leith/apple_google.pdf.

[66] W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel and A. N. Sheth. 'TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones'. In: *Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation*. Osdi'10. Berkeley, CA, United States: USENIX Association, 2010, pp. 393–407.

[67] Y. Song and U. Hengartner. 'PrivacyGuard: A VPN-based Platform to Detect Information Leakage on Android Devices'. In: *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*. Spsm '15. 2015, pp. 15–26. ISBN: 978-1-4503-3819-6. DOI: 10.1145/2808117.2808120.

[68] A. Shuba, A. Markopoulou and Z. Shafiq. 'NoMoAds: Effective and Efficient Cross-App Mobile Ad-Blocking'. In: *Proceedings on Privacy Enhancing Technologies* 2018.4 (Oct. 2018), pp. 125–140. DOI: 10.1515/popets-2018-0035.

[69] C. Han, I. Reyes, A. Elazari, J. Reardon, A. Feal, K. A. Bamberger, S. Egelman and N. Vallina-Rodriguez. 'Do You Get What You Pay For? Comparing The Privacy Behaviors of Free vs. Paid Apps.' In: *The Workshop on Technology and Consumer Protection (ConPro '19)*. 2019. URL: https://www.ieee-security.org/TC/SPW2019/ConPro/papers/han-conpro19.pdf.

[70] I. Reyes, P. Wijesekera, J. Reardon, A. Elazari, A. Razaghpanah, N. Vallina-Rodriguez and S. Egelman. '"Won't Somebody Think of the Children?" Examining COPPA Compliance at Scale'. In: *Proceedings on Privacy Enhancing Technologies* 2018.3 (June 2018), pp. 63–83. DOI: 10.1515/popets-2018-0021.

[71] J. Ren, A. Rao, M. Lindorfer, A. Legout and D. Choffnes. 'ReCon: Revealing and Controlling PII Leaks in Mobile Network Traffic'. In: *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services - MobiSys '16*. MobiSys '16. Singapore, Singapore: ACM Press, 2016, pp. 361–374. ISBN: 978-1-4503-4269-8. DOI: 10.1145/2906388.2906392.

[72] A. Shuba and A. Markopoulou. 'NoMoATS: Towards Automatic Detection of Mobile Tracking'. In: *Proceedings on Privacy Enhancing Technologies* 2020.2 (2020), pp. 45–66. DOI: 10.2478/popets-2020-0017.

[73] E. Okoyomon, N. Samarin, P. Wijesekera, A. Elazari, N. Vallina-Rodriguez, I. Reyes, Á. Feal and S. Egelman. 'On The Ridiculousness of Notice and Consent: Contradictions in App Privacy Policies'. In: *The Workshop on Technology and Consumer Protection (ConPro '19)* (2019). URL: https://www.ieee-security.org/TC/SPW2019/ConPro/papers/okoyomon-conpro19.pdf.

[74] J. Han, Q. Yan, D. Gao, J. Zhou and R. H. Deng. 'Comparing Mobile Privacy Protection through Cross-Platform Applications'. In: *Proceedings 2013 Network and Distributed System Security Symposium*. Network and Distributed System Security Symposium. Internet Society, 2013. URL: https://www.ndss-symposium.org/wp-content/uploads/2017/09/06_2_0.pdf.

[75] M. Egele, C. Kruegel, E. Kirda and G. Vigna. 'PiOS: Detecting Privacy Leaks in iOS Applications'. In: *Proceedings of NDSS 2018*. San Diego, California: The Internet Society, Jan. 2011. URL: https://www.ndss-symposium.org/wp-content/uploads/2017/09/egel.pdf.

[76] N. Viennot, E. Garcia and J. Nieh. 'A Measurement Study of Google Play'. In: *The 2014 ACM International Conference on Measurement and Modeling of Computer Systems*. Sigmetrics '14. 2014, pp. 221–233. ISBN: 978-1-4503-2789-3. DOI: 10.1145/2591971.2592003.

[77] K. Chen, X. Wang, Y. Chen, P. Wang, Y. Lee, X. Wang, B. Ma, A. Wang, Y. Zhang and W. Zou. 'Following Devil's Footprints: Cross-Platform Analysis of Potentially Harmful Libraries on Android and iOS'. In: *2016 IEEE Symposium on Security and Privacy (SP)*. San Jose, CA: Ieee, 2016, pp. 357–376. DOI: 10.1109/sp.2016.29.

[78] J. Guo, M. Zheng, Y. Zhou, H. Wang, L. Wu, X. Luo and K. Ren. *iLibScope: Reliable Third-Party Library Detection for iOS Mobile Apps*. 2022. arXiv: 2207.01837 [cs].

[79] Y. Xiao, Z. Li, Y. Qin, J. Guan, X. Bai, X. Liao and L. Xing. *Lalaine: Measuring and Characterizing Non-Compliance of Apple Privacy Labels at Scale*. 2022. arXiv: 2206.06274 [cs].

[80] Y. Li, D. Chen, T. Li, Y. Agarwal, L. F. Cranor and J. I. Hong. 'Understanding iOS Privacy Nutrition Labels: An Exploratory Large-Scale Analysis of App Store Data'. In: *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI Extended Abstracts '22. New York, NY, USA: Association for Computing Machinery, 2022. DOI: 10.1145/3491101.3519739.

[81] T. Li, K. Reiman, Y. Agarwal, L. F. Cranor and J. I. Hong. 'Understanding Challenges for Developers to Create Accurate Privacy Nutrition Labels'. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Chi '22. New York, NY, USA: Association for Computing Machinery, 2022. DOI: 10.1145/3491102.3502012.

[82] D. Greene and K. Shilton. 'Platform privacies: Governance, collaboration, and the different meanings of "privacy" in iOS and Android development'. In: *New Media & Society* 20.4 (2018), pp. 1640–1657. DOI: 10.1177/1461444817702397.

[83] R. Ó. Fathaigh. 'Mobile Privacy and Business-to-Platform Dependencies: An Analysis of SEC Disclosures'. In: *Journal of Business* (2018).

[84] Datenschutzkonferenz. *Orientierungshilfe Der Aufsichtsbehörden Für Anbieter von Telemedien*. 2019. URL: https://www.datenschutzkonferenz-online.de/media/oh/20190405_oh_tmg.pdf.

[85] R. Balebako, A. Marsh, J. Lin, J. Hong and L. Faith Cranor. 'The Privacy and Security Behaviors of Smartphone App Developers'. In: *Proceedings 2014 Workshop on Usable Security*. Workshop on Usable Security. Internet Society, 2014. ISBN: 978-1-891562-37-2. DOI: 10.14722/usec.2014.23006.

[86] S. Chitkara, N. Gothoskar, S. Harish, J. I. Hong and Y. Agarwal. 'Does this App Really Need My Location?: Context-Aware Privacy Management for Smartphones'. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.3 (2017). DOI: 10.1145/3132029.

[87] I. Hadar, T. Hasson, O. Ayalon, E. Toch, M. Birnhack, S. Sherman and A. Balissa. 'Privacy by designers: software developers' privacy mindset'. In: *Empirical Software Engineering* 23.1 (2018), pp. 259–289. DOI: 10.1007/s10664-017-9517-1.

[88] J. Zhao, G. Wang, C. Dally, P. Slovak, J. Edbrooke-Childs, M. Van Kleek and N. Shadbolt. ''I Make up a Silly Name': Understanding Children's Perception of Privacy Risks Online'. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1–13. DOI: 10.1145/3290605.3300336.

[89] Gegevensbeschermingsautoriteit. *Decision on complaint relating to Transparency & Consent Framework (case number DOS-2019-01377)*. https://www.gegevensbeschermingsautoriteit.be/publications/beslissing-ten-gronde-nr.-21-2022-english.pdf. Feb. 2022.

[90] I. Ajunwa and D. Greene. 'Chapter 3 Platforms at Work: Automated Hiring Platforms and Other New Intermediaries in the Organization of Work'. In: *Research in the Sociology of Work*. Ed. by S. P. Vallas and A. Kovalainen. Vol. 33. Emerald Publishing Limited, 2019, pp. 61–91. DOI: 10.1108/s0277-283320190000033005.

[91] T. Gillespie. 'Regulation of and by Platforms'. In: J. Burgess, A. Marwick and T. Poell. *The SAGE Handbook of Social Media*. SAGE Publications Ltd, 2018, pp. 254–278. DOI: 10.4135/9781473984066.n15.

[92] T. Gillespie. 'Platforms Intervene'. In: *Social Media + Society* 1.1 (2015). DOI: 10.1177/2056305115580479.

[93] T. Gillespie. 'The politics of 'platforms''. In: *New Media & Society* 12.3 (2010), pp. 347–364. DOI: 10.1177/1461444809342738.

[94] N. Helberger. 'How Current Attempts to Regulate Misinformation Amplify Opinion Power'. In: *Digital Journalism* 8.6 (2020), pp. 842–854. DOI: 10.1080/21670811.2020.1773888.

[95] J. van Dijck. *The Culture of Connectivity: A Critical History of Social Media*. Oxford University Press, 2013. 228 pp.

[96] N. Srnicek. *Platform Capitalism*. Theory Redux. Polity, 2016. 171 pp.

[97] A. Holzer and J. Ondrus. 'Mobile Application Market: A Developer's Perspective'. In: *Telematics and Informatics* 28.1 (2011), pp. 22–31. DOI: 10.1016/j.tele.2010.05.006.

[98] B. Bergvall-Kåreborn and D. Howcroft. ''The Future's Bright, the Future's Mobile': A Study of Apple and Google Mobile Application Developers'. In: *Work, Employment and Society* 27.6 (2013), pp. 964–981. DOI: 10.1177/0950017012474709.

[99] Alphabet. *Form 10-k*. 2020. URL: https://abc.xyz/investor/static/pdf/20210203%5C_alphabet%5C_10K.pdf?cache=b44182d.

[100] eMarketer. *Mobile Moves to Majority Share of Google's Worldwide Ad Revenues*. 2016. URL: https://www.emarketer.com/Article/Mobile-Moves-Majority-Share-of-Googles-Worldwide-Ad-Revenues/1014633.

[101] P. Vines, F. Roesner and T. Kohno. 'Exploring ADINT: Using Ad Targeting for Surveillance on a Budget - or - How Alice Can Buy Ads to Track Bob'. In: *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society - WPES '17*. Dallas, Texas, USA: ACM Press, 2017, pp. 153–164. ISBN: 978-1-4503-5175-1. DOI: 10.1145/3139550.3139567.

[102] '50 Ways to Leak Your Data: An Exploration of Apps' Circumvention of the Android Permissions System'. In.

[103] Google. *Device and Network Abuse*. 2021. URL: https://support.google.com/googleplay/android-developer/answer/9888379.

[104] K. D. Martin and P. E. Murphy. 'The Role of Data Privacy in Marketing'. In: *Journal of the Academy of Marketing Science* 45.2 (2017), pp. 135–155. DOI: 10.1007/s11747-016-0495-4.

[105] 9to5mac.com. *Apple rebuffs Facebook criticism, says iOS anti-tracking features are about 'standing up for our users'*. 2020. URL: https://9to5mac.com/2020/12/16/apple-facebook-app-tracking-transparency/.

[106] 9to5mac.com. *Facebook attacks Apple in full-page newspaper ads*. 2020. URL: https://9to5mac.com/2020/12/16/facebook-attacks-apple/.

[107] 9to5mac.com. *Apple versus Facebook on ad-tracking: Harvard sides with Apple*. 2021. URL: https://9to5mac.com/2021/02/05/apple-versus-facebook-harvard/.

[108] J. van Hoboken and R. Ó. Fathaigh. 'Smartphone platforms as privacy regulators'. In: *Computer Law & Security Review* 41 (2021). DOI: 10.1016/j.clsr.2021.105557.

[109] B. Zhou, I. Neamtiu and R. Gupta. 'A Cross-Platform Analysis of Bugs and Bug-Fixing in Open Source Projects: Desktop vs. Android vs. iOS'. In: *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering - EASE '15*. Nanjing, China: ACM Press, 2015. DOI: 10.1145/2745802.2745808.

[110] Free Software Foundation. *More about the App Store GPL Enforcement*. 2010. URL: https://www.fsf.org/blogs/licensing/more-about-the-app-store-gpl-enforcement.

[111] Z. Li, G. Nan and M. Li. 'Effects of Platform Protection in a Duopoly in the Presence of Asymmetric Information and User Security Preference'. In: *SSRN Electronic Journal* (2020). DOI: 10.2139/ssrn.3556488.

[112] P. Roma and D. Ragaglia. 'Revenue Models, in-App Purchase, and the App Performance: Evidence from Apple's App Store and Google Play'. In: *Electronic Commerce Research and Applications* 17 (2016), pp. 173–190. DOI: 10.1016/j.elerap.2016.04.007.

[113] W. Wen and F. Zhu. 'Threat of Platform-owner Entry and Complementor Responses: Evidence from the Mobile App Market'. In: *Strategic Management Journal* 40.9 (2019), pp. 1336–1367. DOI: 10.1002/smj.3031.

[114] P. G. Kelley, L. F. Cranor and N. Sadeh. 'Privacy as Part of the App Decision-Making Process'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. The SIGCHI Conference. ACM Press, 2013. ISBN: 978-1-4503-1899-0. DOI: 10.1145/2470654.2466466.

[115] D. Wetherall. 'Privacy Revelations for Web and Mobile Apps'. In: *HotOS'13: Proceedings of the 13th USENIX conference on Hot topics in operating systems* (2011). URL: https://www.usenix.org/legacy/events/hotos11/tech/final_files/Wetherall.pdf9.

[116] F. Shih, I. Liccardi and D. Weitzner. 'Privacy Tipping Points in Smartphones Privacy Preferences'. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. The 33rd Annual ACM Conference. ACM Press, 2015, pp. 807–816. ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702404.

[117] Court of Justice of the European Union. *Fashion ID GmbH & Co. KG v Verbraucherzentrale NRW e. V.* 2019. URL: http://curia.europa.eu/juris/liste.jsf?num=C-40/17.

[118] Court of Justice of the European Union. *Microsoft v Commission*. 2007. URL: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:62004TJ0201.

[119] *Antitrust: Commission fines Google €4.34 billion for illegal practices regarding Android mobile devices to strengthen dominance of Google's search engine'.*

[120] T. Wu. *The Curse of Bigness: Antitrust in the New Gilded Age*. OCLC: on1029205194. Columbia Global Reports, 2018. 154 pp. ISBN: 978-0-9997454-6-5.

[121] N. R. Lamoreaux. 'The Problem of Bigness: From Standard Oil to Google'. In: *Journal of Economic Perspectives* 33.3 (2019), pp. 94–117. DOI: 10.1257/jep.33.3.94.

[122] L. M. Khan. 'Amazon's Antitrust Paradox'. In: *The Yale Law Journal* (2017).

[123] R. Binns and E. Bietti. 'Dissolving privacy, one merger at a time: Competition, data and third party tracking'. In: *Computer Law & Security Review* 36 (2020), p. 105369. DOI: 10.1016/j.clsr.2019.105369.

[124] O. Lynskey. 'Grappling with "Data Power": Normative Nudges from Data Protection and Privacy'. In: *Theoretical Inquiries in Law* 20.1 (2019), pp. 189–220. DOI: 10.1515/til-2019-0007.

[125] R. Ó Fathaigh and J. van Hoboken. 'European Regulation of Smartphone Ecosystems'. In: *European Data Protection Law Review* 5.4 (2019), pp. 476–491. DOI: 10.21552/edpl/2019/4/6.

[126] Federal Trade Commission. *Mobile Privacy Disclosures–Building Trust Through Transparency*. 2013. URL: https://www.ftc.gov/sites/default/files/documents/reports/mobile-privacy-disclosures-building-trust-through-transparency-federal-trade-commission-staff-report/130201mobileprivacyreport.pdf.

[127] L. A. Bygrave. 'Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements'. In: *Oslo Law Review* 1 (Aug. 2017), pp. 105–120. DOI: 10.18261/issn.2387-3299-2017-02-03.

[128] L. Jasmontaite, I. Kamara, G. Zanfir-Fortuna and S. Leucci. 'Data Protection by Design and by Default: Framing Guiding Principles into Legal Obligations in the GDPR'. In: *European Data Protection Law Review* 4 (June 2018), pp. 168–189. DOI: 10.21552/edpl/2018/2/7.

[129] Authority for Consumers and Markets. *ACM obliges Apple to adjust unreasonable conditions for its App Store*. 2022. URL: https://www.acm.nl/en/publications/acm-obliges-apple-adjust-unreasonable-conditions-its-app-store.

[130] Reuters. *S.Korea targets Apple over new app store regulation*. 2021. URL: https://www.reuters.com/technology/skorea-targets-apple-over-new-app-store-regulation-2021-10-15/.

[131] Sociam. *xray-archiver*. 2018. URL: https://github.com/sociam/xray-archiver.

[132] matlink. *Google Play Downloader via Command line*. URL: https://github.com/matlink/gplaycli.

[133] C. Mallet. *AutoHotkey*. URL: https://www.autohotkey.com/.

[134] D. Orikogbo, M. Büchler and M. Egele. 'CRiOS: Toward Large-Scale iOS Application Analysis'. In: *Proceedings of the 6th Workshop on Security and Privacy in Smartphones and Mobile Devices*. Spsm '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 33–42. DOI: 10.1145/2994459.2994473.

[135] M. Van Kleek, R. Binns, J. Zhao, A. Slack, S. Lee, D. Ottewell and N. Shadbolt. 'X-Ray Refine: Supporting the Exploration and Refinement of Information Exposure Resulting from Smartphone Apps'. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. The 2018 CHI Conference. ACM Press, 2018. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173967.

[136] M. Bokhorst. *NetGuard*. 2021. URL: https://github.com/M66B/NetGuard.

[137] Exodus. *Statistics*. URL: https://reports.exodus-privacy.eu.org/en/trackers/stats/.

[138] O. Geier and D. Herrmann. 'The AppChk Crowd-Sourcing Platform: Which Third Parties are iOS Apps Talking To?' In: June 2021, pp. 228–241. ISBN: 978-3-030-78119-4. DOI: 10.1007/978-3-030-78120-0_15.

[139] European Commission. *EU Leaders' meeting in Sofia: Completing a trusted Digital Single Market for the benefit of all*. 2018. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip%5C_18%5C_3740.

[140] W. Voss. 'European Union Data Privacy Law Reform: General Data Protection Regulation, Privacy Shield, and the Right to Delisting'. In: *The Business Lawyer* 72 (Jan. 2017).

[141] Commission Nationale de l'Informatique et des Libertés. *Décision n° MED 2018-042 du 30 octobre 2018 mettant en demeure la société X*. 2018. URL: https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000037594451/.

[142] Commission Nationale de l'Informatique et des Libertés. *Délibération SAN-2020-012 du 7 décembre 2020*. 2020. URL: https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000042635706.

[143] Information Commissioner's Office. *Update report into adtech and real time bidding*. 2019. URL: https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906-dl191220.pdf.

[144] T. J. McIntyre. 'Regulating the Information Society: Data Protection and Ireland's Internet Industry'. In: *The Oxford Handbook of Irish Politics*. Ed. by D. M. Farrell and N. Hardiman. Oxford University Press, Sept. 2021. ISBN: 978-0-19-882383-4. DOI: 10.1093/oxfordhb/9780198823834.013.39.

[145] E. Massé. *Two Years under GDPR*. Implementation Progress Report. Access Now, 2020. URL: https://www.accessnow.org/cms/assets/uploads/2020/05/Two-Years-Under-GDPR.pdf.

[146] Irish Council for Civil Liberties. *Europe's enforcement paralysis*. 2021. URL: https://www.iccl.ie/digital-data/2021-gdpr-report/.

[147] A. Giannopoulou. 'Algorithmic Systems: The Consent Is in the Detail?' In: *Internet Policy Review* 9.1 (2020). DOI: 10.14763/2020.1.1452.

[148] A. Daly. 'Neo-Liberal Business-As-Usual or Post-Surveillance Capitalism With European Characteristics? The EU's General Data Protection Regulation in a Multi-Polar Internet'. In: *SSRN Electronic Journal* (2020). DOI: 10.2139/ssrn.3655773.

[149] D. Geradin, D. Katsifis and T. Karanikioti. 'GDPR Myopia: How a Well-Intended Regulation Ended up Favoring Google in Ad Tech'. In: *SSRN Electronic Journal* (2020). DOI: 10.2139/ssrn.3598130.

[150] M. S. Gal and O. Aviv. 'The Competitive Effects of the GDPR'. In: *Journal of Competition Law & Economics* 16.3 (2020), pp. 349–391. DOI: 10.1093/joclec/nhaa012.

[151] K. Kollnig. 'Tracking in Apps' Privacy Policies'. In: (2019). arXiv: 2111.07860 [cs].

[152] E. Okoyomon, N. Samarin, P. Wijesekera, A. Elazari, N. Vallina-Rodriguez, I. Reyes, '. Feal and S. Egelman. 'On The Ridiculousness of Notice and Consent: Contradictions in App Privacy Policies'. In: *The Workshop on Technology and Consumer Protection (ConPro '19)*. 2019.

[153] Z. Ma, H. Wang, Y. Guo and X. Chen. 'LibRadar: Fast and Accurate Detection of Third-Party Libraries in Android Apps'. In: *2016 IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C)*. 2016, pp. 653–656. DOI: 10.1145/2889160.2889178.

[154] V. Verouden. *FTC and U.S. DOJ Merger Enforcement Workshop: The role of market shares and market concentration indices in the European Commission's guidelines on the assessment of horizontal mergers under the EC merger regulation*. 2004. URL: https://www.justice.gov/sites/default/files/atr/legacy/2007/08/30/202601.pdf.

[155] S. Englehardt and A. 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. Association for Computing Machinery, 2016, pp. 1388–1401. ISBN: 978-1-4503-4139-4. DOI: 10.1145/2976749.2978313.

[156] Google. *Mobile app marketing trends and mobile landscape*. 2016. URL: https://www.thinkwithgoogle.com/marketing-strategies/app-and-mobile/average-number-of-apps-on-smartphones/.

[157] AudienceProject. *Insights 2019: App & social media use*. 2020. URL: https://www.audienceproject.com/resources/insight-studies/app-social-media-usage/.

[158] Flurry. *iOS 14.5 Opt-in Rate - Daily Updates Since Launch*. 2021. URL: https://www.flurry.com/blog/ios-14-5-opt-in-rate-att-restricted-app-tracking-transparency-worldwide-us-daily-latest-update/.

[159] AppsFlyer. *Initial data indicates ATT opt-in rates are much higher than anticipated – at least 41%*. 2021. URL: https://www.appsflyer.com/blog/trends-insights/att-opt-in-rates-higher/.

[160] Financial Times. *Apple's privacy changes create windfall for its own advertising business*. 2021. URL: https://www.ft.com/content/074b881f-a931-4986-888e-2ac53e286b9d.

[161] Mobile Dev Memo. *ATT advantages Apple's ad network. Here's how to fix that.* 2021. URL: https://mobiledevmemo.com/att-advantages-apples-ad-network-heres-how-to-fix-that/.

[162] European Parliament and Council. *Regulation 2016/679 (General Data Protection Regulation)*. Apr. 2016. URL: http://data.europa.eu/eli/reg/2016/679/oj.

[163] U. S. Congress. *Children's Online Privacy Protection Act*. 1998. URL: https://www.ftc.gov/system/files/2012-31341.pdf.

[164] X. Developers. *Google Play Services will soon delete your advertising ID when you opt out of ad personalization*. 2021. URL: https://www.xda-developers.com/google-play-services-delete-ad-id-opt-out-personalization/.

[165] AdAway. *AdAway*. 2021. URL: https://github.com/AdAway/AdAway.

[166] M. Bokhorst. *XPrivacyLua*. 2021. URL: https://lua.xprivacy.eu/.

[167] Aurora Open Source Software. *Warden : App management utility*. 2021. URL: https://gitlab.com/AuroraOSS/AppWarden.

[168] E. Ferrari-Herrmann. *Trapped in Google's safety net: what modders need to know.* 2021. URL: https://www.nextpit.com/google-safety-net-what-modders-need-to-know.

[169] J. C. Torres. *Google SafetyNet update might be the end for Android rooting, custom ROMs.* 2020. URL: https://www.slashgear.com/google-safetynet-update-might-be-the-end-for-android-rooting-custom-roms-30627121/.

[170] AdGuard. *AdGuard for Android.* 2021. URL: https://adguard.com/en/adguard-android/overview.html.

[171] J. A. Klode. *DNS-Based Host Blocking for Android.* 2021. URL: https://github.com/julian-klode/dns66.

[172] A. Le, J. Varmarken, S. Langhoff, A. Shuba, M. Gjoka and A. Markopoulou. 'AntMonitor: A System for Monitoring from Mobile Devices'. In: *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data.* C2b(1)d '15. Aug. 2015, pp. 15–20. ISBN: 978-1-4503-3539-3. DOI: 10.1145/2787394.2787396.

[173] Google. *Android 7.0 for Developers.* 2016. URL: https://developer.android.com/about/versions/nougat/android-7.0%5C#default%5C_trusted%5C_ca.

[174] Court of Justice of the European Union. *Orange România SA v Autoritatea Națională de Supraveghere a Prelucrării.* 2020. URL: http://curia.europa.eu/juris/document/document.jsf?docid=233544%5C&doclang=EN.

[175] Article 29 Data Protection Working Party. *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC.* 2014. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217%5C_en.pdf.

[176] Article 29 Data Protection Working Party. *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679.* http://ec.europa.eu/newsroom/document.cfm?doc_id=47711. Oct. 2017.

[177] Information Commissioner's Office. *How do we apply legitimate interests in practice?* 2021. URL: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/how-do-we-apply-legitimate-interests-in-practice/.

[178] 'Share First, Ask Later (or Never?) Studying Violations of GDPR's Explicit Consent in Android Apps'. In.

[179] Information Commissioner's Office. *What are the rules on cookies and similar technologies?* 2021. URL: https://ico.org.uk/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies/what-are-the-rules-on-cookies-and-similar-technologies/%5C#rules10.

[180] Information Commissioner's Office. *How do we comply with the cookie rules?* 2021. URL: https://ico.org.uk/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies/how-do-we-comply-with-the-cookie-rules/.

[181] Google. *User Data*. 2021. URL: https://support.google.com/googleplay/android-developer/answer/10144311.

[182] Disconnect.me and Mozilla. *Firefox Blocklist*. URL: https://github.com/mozilla-services/shavar-prod-lists.

[183] Google. *Requesting Consent from European Users*. URL: https://developers.google.com/admob/android/eu-consent.

[184] J. Lawrance, C. Bogart, M. Burnett, R. Bellamy, K. Rector and S. D. Fleming. 'How programmers debug, revisited: An information foraging theory perspective'. In: *IEEE Transactions on Software Engineering* 39.2 (2010), pp. 197–215. DOI: 10.1109/tse.2010.111.

[185] C. Kelleher and M. Ichinco. 'Towards a model of API learning'. In: *2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. 2019, pp. 163–168. DOI: 10.1109/vlhcc.2019.8818850.

[186] Y. Acar, M. Backes, S. Fahl, D. Kim, M. L. Mazurek and C. Stransky. 'You Get Where You're Looking for: The Impact of Information Sources on Code Security'. In: *2016 IEEE Symposium on Security and Privacy (SP)*. 2016, pp. 289–305. DOI: 10.1109/sp.2016.25.

[187] AppsFlyer. *360° Mobile Attribution*. 2021. URL: https://www.appsflyer.com/product/mobile-attribution-for-user-acquisition/.

[188] Google. *Get started with AdMob in your Android project*. 2021. URL: https://firebase.google.com/docs/admob/android/quick-start.

[189] Inmobi. *Android Guidelines: Getting Started with Android SDK Integration*. 2021. URL: https://support.inmobi.com/monetize/android-guidelines/.

[190] AppsFlyer. *Android SDK integration for developers*. 2021. URL: https://support.appsflyer.com/hc/en-us/articles/207032126-Android-SDK-integration-for-developers%5C#integration.

[191] R. Flesch. 'A New Readability Yardstick.' In: *Journal of Applied Psychology* 32.3 (1948), pp. 221–233. DOI: 10.1037/h0057532.

[192] B. G. Edelman and D. Geradin. 'Android and Competition Law: Exploring and Assessing Google's Practices in Mobile'. In: *European Competition Journal* (2016). DOI: 10.1080/17441056.2016.1254483.

[193] Competition and Markets Authority. *Online platforms and digital advertising Market study final report*. Report. 2020. URL: https://assets.publishing.service.gov.uk/media/5efc57ed3a6f4023d242ed56/Final%5C_report%5C_1%5C_July%5C_2020%5C_.pdf.

[194] Court of Justice of the European Union. *Tietosuojavaltuutettu*. 2018. URL: http://curia.europa.eu/juris/document/document.jsf?docid=203822%5C&doclang=EN.

[195] Court of Justice of the European Union. *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH*. 2018. URL: http://curia.europa.eu/juris/liste.jsf?num=C-210/16.

[196] Article 29 Data Protection Working Party. *Opinion 1/2010 on the concepts of "controller" and "processor"*. 2010. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2010/wp169%5C_en.pdf.

[197] European Data Protection Board. *Guidelines 07/2020 on the concepts of controller and processor in the GDPR*. 2020. URL: https://edpb.europa.eu/our-work-tools/documents/public-consultations/2020/guidelines-072020-concepts-controller-and%5C_en.

[198] StatCounter. *Mobile Operating System Market Share in United States Of America - February 2021*. 2021. URL: https://gs.statcounter.com/os-market-share/mobile/united-states-of-america.

[199] Counterpoint Research. *US Monthly Smartphone Sell-Through Highlights Recovery, Device Spec Trends*. 2021. URL: https://www.counterpointresearch.com/us-monthly-smartphone-sell-highlights-recovery/.

[200] US House of Representatives Judiciary Subcommittee on Antitrust. *Investigation of Competition in Digital Markets*. 2020. URL: https://judiciary.house.gov/uploadedfiles/competition%5C_in%5C_digital%5C_markets.pdf?utm%5C_campaign=4493-519.

[201] US Department of Justice. *Complaint, United States v. Google LLC, No. 1:20-cv-03010*. 2020. URL: https://www.justice.gov/opa/press-release/file/1328941/download.

[202] European Commission. *Antitrust: Commission opens investigations into Apple's App Store rules*. 2020. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip%5C_20%5C_1073.

[203] Z. Tang, K. Tang, M. Xue, Y. Tian, S. Chen, M. Ikram, T. Wang and H. Zhu. 'iOS, Your OS, Everybody's OS: Vetting and Analyzing Network Services of iOS Applications'. In: *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, 2020, pp. 2415–2432. ISBN: 978-1-939133-17-5. URL: https://www.usenix.org/conference/usenixsecurity20/presentation/tang.

[204] S. Zimmeck, Z. Wang, L. Zou, R. Iyengar, B. Liu, F. Schaub, S. Wilson, N. Sadeh, S. M. Bellovin and J. Reidenberg. 'Automated Analysis of Privacy Requirements for Mobile Apps'. In: *NDSS Symposium 2017*. Feb. 2017. DOI: 10.14722/ndss.2017.23034.

[205] A. P. Felt, E. Chin, S. Hanna, D. Song and D. Wagner. 'Android Permissions Demystified'. In: *Proceedings of the 18th ACM Conference on Computer and Communications Security - CCS '11*. The 18th ACM Conference. ACM Press, 2011. ISBN: 978-1-4503-0948-6. DOI: 10.1145/2046707.2046779.

[206] J. Jeon, K. K. Micinski, J. A. Vaughan, A. Fogel, N. Reddy, J. S. Foster and T. Millstein. 'Dr. Android and Mr. Hide: Fine-Grained Permissions in Android Applications'. In: *Proceedings of the Second ACM Workshop on Security and Privacy in Smartphones and Mobile Devices - SPSM '12*. The Second ACM Workshop. Raleigh, North Carolina, USA: ACM Press, 2012. ISBN: 978-1-4503-1666-8. DOI: 10.1145/2381934.2381938.

[207] B. Liu, M. S. Andersen, F. Schaub, H. Almuhimedi, S. Zhang, N. Sadeh, A. Acquisti and Y. Agarwal. 'Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions'. In: *Proceedings of the Fifteenth Symposium on Usable Privacy and Security* (2016).

[208] J. Lin, B. Liu, N. Sadeh and J. I. Hong. 'Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings'. In: *Proceedings of the Fifteenth Symposium on Usable Privacy and Security* (2014).

[209] P. Wijesekera, A. Baokar, L. Tsai, J. Reardon, S. Egelman, D. Wagner and K. Beznosov. 'The Feasibility of Dynamically Granted Permissions: Aligning Mobile Privacy with User Preferences'. In: *2017 IEEE Symposium on Security and Privacy (SP)*. 2017 IEEE Symposium on Security and Privacy (SP). Ieee, 2017, pp. 1077–1093. ISBN: 978-1-5090-5533-3. DOI: 10.1109/sp.2017.51.

[210] A. P. Felt, K. Greenwood and D. Wagner. 'The Effectiveness of Application Permissions'. In: *Proceedings of the 2Nd USENIX Conference on Web Application Development*. WebApps'11. 2011.

[211] D. Barrera, H. G. Kayacik, P. C. van Oorschot and A. Somayaji. 'A Methodology for Empirical Analysis of Permission-based Security Models and Its Application to Android'. In: *Proceedings of the 17th ACM Conference on Computer and Communications Security*. The 17th ACM Conference. Ccs '10. New York, NY, USA: ACM Press, 2010, pp. 73–84. ISBN: 978-1-4503-0245-6. DOI: 10.1145/1866307.1866317.

[212] A. Shuba, A. Le, E. Alimpertis, M. Gjoka and A. Markopoulou. 'AntMonitor: A System for On-Device Mobile Network Monitoring and its Applications'. In: *arXiv preprint arXiv:1611.04268* (2016).

[213] CocoaPods. *Master Repo*. URL: https://github.com/CocoaPods/Specs.

[214] M. Backes, S. Bugiel and E. Derr. 'Reliable Third-Party Library Detection in Android and its Security Applications'. In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. Ccs '16. New York, NY, USA: Acm, 2016, pp. 356–367. ISBN: 978-1-4503-4139-4. DOI: 10.1145/2976749.2978333.

[215] E. Derr, S. Bugiel, S. Fahl, Y. Acar and M. Backes. 'Keep Me Updated: An Empirical Study of Third-Party Library Updatability on Android'. In: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. Ccs '17. New York, NY, USA: Acm, 2017, pp. 2187–2200. ISBN: 978-1-4503-4946-8. DOI: 10.1145/3133956.3134059.

[216] Google. *Advertising ID - Play Console Help*. URL: https://support.google.com/googleplay/android-developer/answer/6048248.

[217] Privacy International. *How Apps on Android Share Data with Facebook*. 2018. URL: https://privacyinternational.org/campaigns/investigating-apps-interactions-facebook-android.

[218] Yuanchun Li, Ziyue Yang, Yao Guo and Xiangqun Chen. 'DroidBot: A Lightweight UI-Guided Test Input Generator for Android'. In: *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*. Buenos Aires: Ieee, May 2017, pp. 23–26. ISBN: 978-1-5386-1589-8. DOI: 10.1109/icse-c.2017.8.

[219] Bundeskartellamt. *Proceeding against Google based on new rules for large digital players*. 2021. URL: https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2021/25%5C_05%5C_2021%5C_Google%5C_19a.html.

[220] N. Shuyo and M. Danilák. *langdetect*. URL: https://pypi.org/project/langdetect/.

[221] Court of Justice of the European Union. *Breyer v Germany*. 2016. URL: https://curia.europa.eu/juris/document/document.jsf?docid=184668%5C&doclang=EN.

[222] Financial Times. *China's tech giants test way around Apple's new privacy rules*. 2021. URL: https://www.ft.com/content/520ccdae-202f-45f9-a516-5cbe08361c34.

[223] R. D. Binns, D. Millard and L. Harris. 'Data Havens, or Privacy sans Frontières? A Study of International Personal Data Transfers'. In: *Proceedings of the 2014 ACM Conference on Web Science*. WebSci '14. New York, NY, USA: Association for Computing Machinery, 2014, pp. 273–274. DOI: 10.1145/2615569.2615650.

[224] Information Commissioner's Office. *Age appropriate design: a code of practice for online services*. 2020. URL: https://ico.org.uk/for-organisations/guide-to-data-protection/ico-codes-of-practice/age-appropriate-design-a-code-of-practice-for-online-services/.

[225] Apple. *Updates to the App Store Review Guidelines*. 2019. URL: https://developer.apple.com/news/?id=06032019j.

[226] Apple. *App Store Review Guidelines*. URL: https://developer.apple.com/app-store/review/guidelines/.

[227] Google. *Developer Content Policy*. URL: https://play.google.com/about/developer-content-policy/.

[228] StatCounter. *Mobile & Tablet Android Version Market Share United Kingdom*. 2020. URL: https://gs.statcounter.com/android-version-market-share/mobile-tablet/united-kingdom/%5C#monthly-201912-202003.

[229] K. Kollnig and R. Binns. *The Cost of the GDPR for Apps? Nearly Impossible to Study without Platform Data*. 2022. arXiv: 2206.09734 [cs].

[230] Apple. *Expanded Protections for Children*. 2021. URL: https://www.apple.com/child-safety/.

[231] H. Abelson, R. Anderson, S. M. Bellovin, J. Benaloh, M. Blaze, J. Callas, W. Diffie, S. Landau, P. G. Neumann, R. L. Rivest, J. I. Schiller, B. Schneier, V. Teague and C. Troncoso. 'Bugs in Our Pockets: The Risks of Client-Side Scanning'. In: (2021). arXiv: 2110.07450 [cs].

[232] International Association of Privacy Professionals. *Apple's ATT rollout presents uncertain path for adtech*. 2021. URL: https://iapp.org/news/a/apples-att-rollout-presents-uncertain-path-for-adtech/.

[233] tinuiti. *What is an IDFA (and How Will Apple's iOS 14 Update Impact Advertisers)?* 2021. URL: https://tinuiti.com/blog/data-privacy/apple-ios-idfa-guide/.

[234] Financial Times. *Snap, Facebook, Twitter and YouTube lose nearly $10bn after iPhone privacy changes*. 2021. URL: https://www.ft.com/content/4c19e387-ee1a-41d8-8dd2-bc6c302ee58e.

[235] Financial Times. *Alphabet and Microsoft smash estimates with $110bn revenue haul*. 2021. URL: https://www.ft.com/content/273aeecb-57a8-40f8-a2ba-8a21a635b289.

[236] R. Kesler. 'The Impact of Apple's App Tracking Transparency on App Monetization'. In: *Work in Progress* (2022).

[237] Mobile Dev Memo. *Why isn't Apple policing mobile ads fingerprinting?* 2021. URL: https://mobiledevmemo.com/why-isnt-apple-policing-mobile-ads-fingerprinting/.

[238] Lockdown Privacy. *Study: Effectiveness of Apple's App Tracking Transparency*. 2021. URL: https://blog.lockdownprivacy.com/2021/09/22/study-effectiveness-of-apples-app-tracking-transparency.html.

[239] Washington Post. *When you 'Ask app not to track,' some iPhone apps keep snooping anyway*. 2021. URL: https://www.washingtonpost.com/technology/2021/09/23/iphone-tracking/.

[240] P. G. Kelley, J. Bresee, L. F. Cranor and R. W. Reeder. 'A "Nutrition Label" for Privacy'. In: *Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS '09*. The 5th Symposium. ACM Press, 2009. ISBN: 978-1-60558-736-3. DOI: 10.1145/1572532.1572538.

[241] P. G. Kelley, L. Cesca, J. Bresee and L. F. 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. New York, NY, USA: Association for Computing Machinery, 2010, pp. 1573–1582. DOI: 10.1145/1753326.1753561.

[242] Washington Post. *I checked Apple's new privacy 'nutrition labels.' Many were false.* 2021. URL: https://www.washingtonpost.com/technology/2021/01/29/apple-privacy-nutrition-label/.

[243] Frida. *Frida: A world-class dynamic instrumentation framework*. URL: https://frida.re.

[244] The Verge. *Apple to finally stop accepting apps that use outdated UDID device identifier on May 1st*. 2013. URL: https://www.theverge.com/2013/3/21/4133288/apple-to-finally-stop-accepting-apps-that-use-outdated-udid-device-identifier-may-1st.

[245] Financial Times. *Apple reaches quiet truce over iPhone privacy changes*. 2021. URL: https://www.ft.com/content/69396795-f6e1-4624-95d8-121e4e5d7839.

[246] W3C Working Group. *Tracking Compliance and Scope*. 2019. URL: https://www.w3.org/TR/tracking-compliance/%5C#tracking.

[247] Commission Nationale de l'Informatique et des Libertés. *Advertising ID: APPLE DISTRIBUTION INTERNATIONAL fined 8 million euros*. 2023. URL: https://www.cnil.fr/en/advertising-id-apple-distribution-international-fined-8-million-euros.

[248] Competition and Markets Authority. *Investigation into Google's 'Privacy Sandbox' browser changes*. 2021. URL: https://www.gov.uk/cma-cases/investigation-into-googles-privacy-sandbox-browser-changes.

[249] M. Veale, M. Nouwens and C. Santos. 'Impossible Asks: Can the Transparency and Consent Framework Ever Authorise Real-Time Bidding After the Belgian DPA Decision?' In: *Technology and Regulation* 2022 (Feb. 2022), pp. 12–22. DOI: 10.26116/techreg.2022.002.

[250] Austrian Data Protection Authority. *Partial Decision D155.027, 2021-0.586.257*. 2021. URL: https://noyb.eu/sites/default/files/2022-01/E-DSB%5C%20-%5C%20Google%5C%20Analytics_EN_bk.pdf.

[251] Commission Nationale de l'Informatique et des Libertés. *Use of Google Analytics and data transfers to the United States: the CNIL orders a website manager/operator to comply*. 2022. URL: https://www.cnil.fr/en/use-google-analytics-and-data-transfers-united-states-cnil-orders-website-manageroperator-comply.

[252] European Data Protection Supervisor. *Decision of the European Data Protection Supervisor in complaint case 2020-1013 submitted by Members of the Parliament against the European Parliament*. 2022. URL: https://noyb.eu/sites/default/files/2022-01/Case%5C%202020-1013%5C%20-%5C%20EDPS%5C%20Decision_bk.pdf.

[253] M. Veale, R. Binns and J. Ausloos. 'When Data Protection by Design and Data Subject Rights Clash'. In: *International Data Privacy Law* 8.2 (Apr. 2018), pp. 105–123. DOI: 10.1093/idpl/ipy002.

[254] Google. *Understanding PII in Google's contracts and policies*. 2021. URL: https://support.google.com/analytics/answer/7686480.

[255] Court of Justice of the European Union. *Patrick Breyer v Bundesrepublik Deutschland*. 2016. URL: https://curia.europa.eu/juris/document/document.jsf?docid=184668.

[256] C. Norval, H. Janssen, J. Cobbe and J. Singh. 'RECLAIMING Data: Overcoming App Identification Barriers for Exercising Data Protection Rights'. In: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. UbiComp '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 921–930. ISBN: 978-1-4503-5966-5. DOI: 10.1145/3267305.3274153.

[257] S. Wachter. 'Data Protection in the Age of Big Data'. In: *Nature Electronics* 2 (Jan. 2019). DOI: 10.1038/s41928-018-0193-y.

[258] Euractiv. *Apple slams sideloading provisions in the DMA*. 2021. URL: https://www.euractiv.com/section/digital/news/apple-slams-sideloading-provisions-in-the-dma/.

[259] F. J Zuiderveen Borgesius, S. Kruikemeier, S. C Boerman and N. Helberger. 'Tracking Walls, Take-It-Or-Leave-It Choices, the GDPR, and the ePrivacy Regulation'. In: *European Data Protection Law Review* 3.3 (Jan. 2017), pp. 353–368. DOI: 10.21552/edpl/2017/3/9.

[260] M. Veale and F. Zuiderveen Borgesius. 'Adtech and Real-Time Bidding under European Data Protection Law'. In: *German Law Journal* 23.2 (2022), pp. 226–256. DOI: 10.1017/glj.2022.18.

[261] Datenschutzkonferenz. *Orientierungshilfe Der Aufsichtsbehörden Für Anbieter von Telemedien*. 2021. URL: https://datenschutzkonferenz-online.de/media/oh/20211220_oh_telemedien.pdf.

[262] Norwegian Consumer Council. *Out of Control: How Consumers Are Exploited by the Online Advertising Industry*. Tech. rep. 2020. URL: https://fil.forbrukerradet.no/wp-content/uploads/2020/01/2020-01-14-out-of-control-final-version.pdf.

[263] European Data Protection Board. *Overview on resources made available by Member States to the Data Protection Supervisory Authorities*. 2022. URL: https://edpb.europa.eu/system/files/2022-09/edpb_overviewresourcesmade_availablebymemberstatestosas2022%5C_en.pdf.

[264] K. Kollnig and N. Shadbolt. *Ready for the EU Digital Services Act? How Decisions by Apple and by Google Impede App Privacy*. SSRN Scholarly Paper. Rochester, NY, Jan. 2023. DOI: 10.2139/ssrn.4343640.